

STIC-Biotech/ChemLib

192484

From: Li, Ruixiang
Sent: Friday, June 09, 2006 2:23 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search of Application No.10/644,084

Please do a standard search on:

- (i). SEQ ID NO: 2 against amino acid databases (excluding pending databases).
(ii). SEQ ID NOS: 1 and 2 against nucleic acid databases (excluding pending databases).

Thank you very much!

Ruixiang Li
GAU 1646
REM 4D75
Mail Box 4C70
(571) 272-0875

78733

URFE

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: _____
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

THIS PAGE LEFT BLANK

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model
Run on: June 13, 2006, 20:46:55 ; Search time 11414 Seconds
(without alignments)
13188.614 Million cell updates/sec

Title: US-10-644-084-1
Perfect score: 2692
Sequence: 1 cgtgagagtgacagagc.....ttgattatggagtaatgggg 2692

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 27959665780:residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: gb_est1:*
 - 2: gb_est2:*
 - 3: gb_est3:*
 - 4: gb_est4:*
 - 5: gb_est5:*
 - 6: gb_est6:*
 - 7: gb_est7:*
 - 8: gb_est8:*
 - 9: gb_est9:*
 - 10: gb_est10:*
 - 11: gb_gsa1:*
 - 12: gb_gsa2:*
 - 13: gb_gsa3:*
 - 14: gb_gsa4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2676	99.4	3185	AK043865	AK043865 Mus muscu
2	2676	99.4	3413	AK031356	AK031356 Mus muscu
3	1985.4	73.8	2245	AK077112	AK077112 Mus muscu
4	1973	73.3	2371	AK049080	AK049080 Mus muscu
5	1832	68.1	1845	AY412493	AY412493 Mus muscu
6	1701.4	63.2	3154	BC035580	BC035580 Homo sapi
7	1701.4	63.2	5658	HS0801317	HS0801317 Homo sapi
8	1416.6	52.6	2111	CR613292	CR613292 full-leng
9	1363.8	50.7	1845	AY412491	AY412491 Homo sapi
10	1357.4	50.4	1845	AY412492	AY412492 Homo sapi
11	876.4	32.6	878	AY412493	AY412493 Pan trogl
12	812	30.2	840	CF949831	CF949831 UI-M-HU0
13	810	30.1	834	CF949831	CF949831 UI-M-HU0
14	802	29.8	804	CF949831	CF949831 UI-M-HU0
15	798	29.6	798	CF949831	CF949831 UI-M-HU0
16	789	29.3	822	CF949831	CF949831 UI-M-HU0
17	768.4	28.5	771	CF949831	CF949831 UI-M-HU0
18	761	28.3	933	CF949831	CF949831 UI-M-HU0
19	753.2	28.0	771	CF949831	CF949831 UI-M-HU0

20 750.2 27.9 770 5 CF729394
21 749 27.8 760 8 CN528686
22 748.2 27.8 845 3 BQ763157
23 739.2 27.5 746 8 CN528831
24 721.4 26.8 723 5 CF737957
25 710.4 26.4 712 5 CF737957
26 690 25.6 736 8 CX226608
27 679 25.2 679 8 CX226608
28 677.8 25.2 683 8 CX226608
29 675 25.1 687 9 CX568194
30 658 24.4 903 10 DV893190
31 653.2 24.3 1005 1 AL582610
32 643 23.9 643 2 BG085173
33 638 23.7 640 5 CF737822
34 633 23.5 633 8 CV117234
35 623.4 23.2 638 8 CN120771
36 622.6 23.1 954 1 AL561392
37 622.2 23.1 635 5 CF899025
38 616.4 22.9 860 10 DT831879
39 614.4 22.8 633 1 AV279314
40 614.2 22.8 805 2 BG242849
41 614.2 22.8 828 10 DV803752
42 612.2 22.7 658 7 BB614736
43 606.8 22.5 623 4 CA548100
44 601.2 22.3 752 7 BF539334
45 600 22.3 600 2 BG804175

ALIGNMENTS

RESULT 1
AK043865
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AK043865 3185 bp mRNA linear HTC 02-SEP-2005
Mus musculus 10 days neonate cortex cDNA, RIKEN full-length
enriched library, clone: A830043F14 product: HYPOTHETICAL 71.0 KDA
PROTEIN homolog [Mus musculus], full insert sequence.
AK043865.1 GI:26335971
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
11042159

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuda, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE

The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection

THIS PAGE LEFT BLANK

THIS PAGE LEFT BLANK

3a

QY 901 CGCGAGCTGAAGAGTCTCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTCTCTCC 960
Db 1210 CGCGAGCTGAAGAGTCTCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTCTCTCC 1269
QY 961 TCAGAAAGAGAGCCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db 1270 TCAGAAAGAGAGCCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1329
QY 1021 TATAGAAGATGACTCTGGGAACTGAGCAGAGACAGAGAGAGAGAGAGAGAGAGAGAG 1080
Db 1330 TATAGAAGATGACTCTGGGAACTGAGCAGAGACAGAGAGAGAGAGAGAGAGAGAGAG 1389
QY 1081 TGTGAGAGAGCAGTGAACAAACAGCATCAGAAACAGTGAAGAAATTTGAAAAGTCACTG 1140
Db 1390 TGTGAGAGAGCAGTGAACAAACAGCATCAGAAACAGTGAAGAAATTTGAAAAGTCACTG 1449
QY 1141 AGAAAGTTCGATAACCAAGCTTCGAAGGTACACTGAGAGGGCTTAAATGAGGAGAGAG 1200
Db 1450 AGAAAGTTCGATAACCAAGCTTCGAAGGTACACTGAGAGGGCTTAAATGAGGAGAGAG 1509
QY 1201 CATCTCAGCAGAGACCATGAGCAAGAGACTGAGAACTGAGAACTGAGAGCTGGAATGAGCGGTG 1260
Db 1510 CATCTCAGCAGAGACCATGAGCAAGAGACTGAGAACTGAGAACTGAGAGCTGGAATGAGCGGTG 1569
QY 1261 TAAAGAGATGATCAAGGCTCAGCAGAGCTCTTACAGCAGAGAGCTGGCCACCACTGAGTGA 1320
Db 1570 TAAAGAGATGATCAAGGCTCAGCAGAGCTCTTACAGCAGAGAGCTGGCCACCACTGAGTGA 1626
QY 1321 TGATGACACCACTCACTCTGCGAGACTGTTTACCTGCGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1627 TGATGACACCACTCACTCTGCGAGACTGTTTACCTGCGAGAGAGAGAGAGAGAGAGAGAG 1686
QY 1381 AGAAGAGTGGAGCCCTTTTAAAGAGCAAAAGAAATTTGAGAGAGAGAGAGAGAGAGAG 1440
Db 1687 AGAAGAGTGGAGCCCTTTTAAAGAGCAAAAGAAATTTGAGAGAGAGAGAGAGAGAGAG 1746
QY 1441 TACAGAGCTGCCATTCGATTTGGGCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1747 TACAGAGCTGCCATTCGATTTGGGCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1806
QY 1501 CTGGGTAAGCAGAGAGTTTAAACATGACGAACTTTGACCAAGAGAGAGAGAGAGAGAGAG 1560
Db 1807 CTGGGTAAGCAGAGAGTTTAAACATGACGAACTTTGACCAAGAGAGAGAGAGAGAGAGAG 1866
QY 1561 GAAACTTTTACAGTCTCTCAGAGAGTCTGATCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
Db 1867 GAAACTTTTACAGTCTCTCAGAGAGTCTGATCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1926
QY 1621 GCCACGCAAAAGAGCTACACAGTGTGGCTTAATGGGGTGCAGCTTGCACATCAAAACT 1680
Db 1927 GCCACGCAAAAGAGCTACACAGTGTGGCTTAATGGGGTGCAGCTTGCACATCAAAACT 1986
QY 1681 GACTAAATCTCTCTGCTCACTTCTTACAGTCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Db 1987 GACTAAATCTCTCTGCTCACTTCTTACAGTCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAG 2046
QY 1741 GTCTGAACACAGTTCCATCAGTGTGCTGAATATTAACCTCTGAGAGAGAGAGAGAGAGAG 1800
Db 2047 GTCTGAACACAGTTCCATCAGTGTGCTGAATATTAACCTCTGAGAGAGAGAGAGAGAGAG 2106
QY 1801 GGTGCAAG 1860
Db 2107 GGTGCAAG 2166
QY 1861 GGGGTGCTACAGCGAGATGCTCTGCGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db 2167 GGGGTGCTACAGCGAGATGCTCTGCGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2226
QY 1921 ACCTTAAATGTGGGCTGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
Db 2227 ACCTTAAATGTGGGCTGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2286
QY 1981 GTGTAGCATAAAGTCAGTCTGCTAACTTAAGATGCTCAGAGTGTGTTGTTGAGCTTCGC 2040

Db 2287 GTGTAGCATAAAGTCAGTCTGCTAACTTAAGATGCTCAGAGTGTGTTGTTGAGCTTCGC 2346
QY 2041 TGTCTTCCCCCAAGAGAGCTGAATGCTTAAGTACTTAAAGGATGCAAGAGCTTTGTTGT 2100
Db 2347 TGTCTTCCCCCAAGAGAGCTGAATGCTTAAGTACTTAAAGGATGCAAGAGCTTTGTTGT 2406
QY 2101 GTGTAGTAAACAG 2160
Db 2407 GTGTAGTAAACAG 2466
QY 2161 ACAG 2220
Db 2467 ACAG 2526
QY 2221 TCTCTGTAACATTAATAATATTCCTCCCAAGTGTGCTGAGAGAGAGAGAGAGAGAGAGAG 2280
Db 2527 TCTCTGTAACATTAATAATATTCCTCCCAAGTGTGCTGAGAGAGAGAGAGAGAGAGAGAG 2586
QY 2281 TGAATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
Db 2587 TGAATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2646
QY 2341 TTTGGGATTAACCTCTTCTCATATGTTTGTGCTGTACATAAAATATACATGATGATGAT 2400
Db 2647 TTTGGGATTAACCTCTTCTCATATGTTTGTGCTGTACATAAAATATACATGATGATGAT 2706
QY 2401 GAGGCTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAG 2460
Db 2707 GAGGCTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAG 2766
QY 2461 AATGTTTTCAGTGGCATGCTGTTTACAGAGAGTGTAGGAGAGAGAGAGAGAGAGAGAGAG 2520
Db 2767 AATGTTTTCAGTGGCATGCTGTTTACAGAGAGTGTAGGAGAGAGAGAGAGAGAGAGAGAG 2826
QY 2521 CTTTATTTTAAAGGCACTCTGGGCAATTAAGGCAATTAAGGCAATTAAGGCAATTAAGG 2580
Db 2827 CTTTATTTTAAAGGCACTCTGGGCAATTAAGGCAATTAAGGCAATTAAGGCAATTAAGG 2886
QY 2581 ATAACTAGAGTCTTAAAGGCACTCTGGGCAATTAAGGCAATTAAGGCAATTAAGGCAAT 2640
Db 2887 ATAACTAGAGTCTTAAAGGCACTCTGGGCAATTAAGGCAATTAAGGCAATTAAGGCAAT 2946
QY 2641 ATATGTGAATGTTATTAATTTCTAAGAGGATATTTGATTTGATTTGATTTGATTTGATTT 2692
Db 2947 ATATGTGAATGTTATTAATTTCTAAGAGGATATTTGATTTGATTTGATTTGATTTGATTT 2998

AK031356 3413 bp mRNA linear HTC 02-SEP-2005
Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
enriched library, clone:6030411i12 product:hypothetical protein,
full insert sequence.
AK031356
AK031356.1 GI:26327238
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
10349636
2
Carninci, P., Shibata, Y., Hayashizaki, Y., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

THIS PAGE LEFT BLANK

THIS PAGE LEFT BLANK

THIS PAGE LEFT BLANK.

Db 2773 ATAAGTGAAGTACTAAGCTTACATTTTGTGAGATTTTAAAGCATTTGATTTATTTAT 2832
 Qy 2641 ATATGTGATGTTATTAATTTCTTAAGAGCAATATTTGATTTATGAGTAATGGG 2692
 Db 2833 ATATGTGATGTTATTAATTTCTTAAGAGCAATATTTGATTTATGAGTAATGGG 2884

RESULT 3
 BC021749
 LOCUS
 DEFINITION
 ACCSSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BC021749 3425 bp mRNA linear ROD 18-JUL-2005
 Mus musculus synovial sarcoma, X breakpoint 2 interacting protein,
 mRNA (CDNA clone MGC:25823 IMAGE:4165430), complete cds.
 BC021749
 MGC.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 3425)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
 Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
 Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
 Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,
 Diatchenko, L., Marusik, K., Farmer, A.A., Rubin, G.M., Hong, L.,
 Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,
 Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
 Carninci, P., Prange, C., Raja, S.S., Iqbal, N.A., Peters, G.J.,
 Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J.,
 McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
 Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
 Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
 Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,
 Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
 Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywicki, M.I., Skalek, U., Smal, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Mammalian Gene Collection Program Team
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)
 12477932
 2 (bases 1 to 3425)

NH MGC Project
 Direct Submission
 Submitted (18-JAN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegod, H.,
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 30 Row: m Column: 16.
 Location/Qualifiers
 1..3425
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"

FEATURES
 source
 1..3425
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"

gene

CDS

ORIGIN

Query Match 99.9%; Score 2688.8; DB 6; Length 3425;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2690; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 CGTAGAGAGTGCAGAGAGCTGTTGTAAGCGTCGACGAGCTGAGCGCGCTCTCCAGGTAT 60
 Db 161 CGTAGAGAGTGCAGAGAGCTGTTGTAAGCGTCGACGAGCTGAGCGCGCTCTCCAGGTAT 220
 Qy 61 CTGGCTCTGGAACTTGTCTATGGAGATTTGGATGACTGTGACAGATTCAGTCTCTGTGTAC 120
 Db 221 CTGGCTCTGGAACTTGTCTATGGAGATTTGGATGACTGTGACAGATTCAGTCTCTGTGTAC 280
 Qy 121 AGAAACAAATCTCTCAATATACCTCAGAAACAAAGATGTCTCCGTCAGATTTGTA 180
 Db 281 AGAAACAAATCTCTCAATATACCTCAGAAACAAAGATGTCTCCGTCAGATTTGTA 340
 Qy 181 CTCCAGCAGAGTCTGTGTCTCTTACGACCTTTATCCAAACAAAGATGTCTCCGTCAG 240
 Db 341 CTCCAGCAGAGTCTGTGTCTCTTACGACCTTTATCCAAACAAAGATGTCTCCGTCAG 400
 Qy 241 TGTCTTCTGCAGAGAGAGCAATTTGAAACAAAGATTTTCTATCTTGTATCAGGAGCTGAC 300
 Db 401 TGTCTTCTGCAGAGAGAGCAATTTGAAACAAAGATTTTCTATCTTGTATCAGGAGCTGAC 460
 Qy 301 CACCTTCGGGTTCTTCTTCTGTATGAAGATTCCTAAAGATGAAGAGGCAAGAGAGATT 360
 Db 461 CACCTTCGGGTTCTTCTTCTGTATGAAGATTCCTAAAGATGAAGAGGCAAGAGAGATT 520
 Qy 361 AAATATAGTCGCTGTTCTGAACCTGTATGAACGAGCTGCTGCTGCTTCAGCGGAGAACCT 420
 Db 521 AAATATAGTCGCTGTTCTGAACCTGTATGAACGAGCTGCTGCTGCTTCAGCGGAGAACCT 580
 Qy 421 GCTGGCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGGGAGTGAATGGACCACT 480
 Db 581 GCTGGCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGGGAGTGAATGGACCACT 640
 Qy 481 GCAGAGCTGCTACGGCAAACTTTAAGAGCAGTTGAAACAGTCCAGCGGAGATGATCGG 540
 Db 641 GCAGAGCTGCTACGGCAAACTTTAAGAGCAGTTGAAACAGTCCAGCGGAGATGATCGG 700

THIS PAGE LEFT BLANK

THIS PAGE LEFT BLANK

301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320
1172 GAAGAGCAGAGGAGCGGACAGGACCTGTTCTATCTCCGATATAGAGATGACTCTGGG 1231
321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluLeuThr 340
1232 GAACCTGACAGAGACAGCGTGTGGGCGCTTCTCTGTGACACTGTGACAGAGCAGCTGACA 1291
341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
1292 AACAGCATCAGGAACAGTGGAGAAATTTTGAAGAAGTGTATGAGAAACTCGATACCA 1351
361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
1352 GCTTCGAAGGTACACTCAGAGGCGCTTATGAGGAGGAGCTCATCTCCAGCAAGACCAT 1411
381 GluGlnGluThrGluLysLeuGluLeuGluLeuGluArgCysLysGluMetIleLysAla 400
1412 GAGCAAGAGACTGAGAAACTGGAGCTGGAGATTGAGCGGTGTAAAGAGATGATCAAGCT 1471
401 GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
1472 CAGCAGCAGCTTACAGCAGCAGCTGGCCACCACGCTGTGATGATGACACCACTCACTG 1531
421 LeuArgAspCysThrLeuLeuGluGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440
1532 TTGCGAGAGCTGTACTTCTGCGAGAAAGGAAAGGAGCCCTTAAAGAGAGTGGACCTTTT 1591
441 LysGluGlnLysLysAsnPheGluArgGluArgGserPheThrGluAlaAlaIleArg 460
1592 AAAGAGCAAAAAAGAAATTTTTCAGAGAGAAAGCGCAAGCTTTACAGAAAGCTGCCATTGCA 1651
461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPhe 480
1652 TTGGGCTTGGAGAGAAAGCGCTTGAAGAAGAGCGAGCGAGCTGGGTAAAGCAGCAGTT 1711
481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
1712 TTAACATACCAACTTTGACCAACCAACTCAGAAATGTGAATCTTTCAGTCCCTTC 1771
501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
1772 TCAGGAAGTTCTGATCCAGACAATCTTATGATCCACTCAGCGCCACGCAAGAAAGCTA 1831
521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
1832 CACAGTGTGGCTTAATGGGTGGCAGCTTGCACATCAAACTGACTAAATCTCTCTCTGCC 1891
541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerIle 560
1892 TCACCTTCTACTTCAGACTTTGCCAGACACATTCATGTGTCTGACACAGTTCCTCATC 1951
561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580
1952 AGTGTCTGATATAACTCTGAGAAAGTAAACCAAGTGGAGTTCAGAGAAAGACGACG 2011
581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGlyCys 600
2012 GATCAGAAGTGGAGCGTGCAGTGCAGCGCCAGCTCGCGGAGGGGTGTACAGCGGATGC 2071
601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
2072 TCCTCGGCTTTCAGGAGCGCTACCGGGACCGAGATGACTTACT 2116

RESULT 3
BC021749
LOCUS
DEFINITION
Mus musculus synovial sarcoma, X breakpoint 2 interacting protein,
mRNA (cDNA clone MGC:25823 IMAGE:4165430), complete cds.
ACCESSION
BC021749
VERSION
BC021749.1 GI:18256805
KEYWORDS
MGC.

SOURCE
ORGANISM

Mus musculus (house mouse)

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS

1 (bases 1 to 3425)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Sapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schect, T.E., Brownstein, M.J., Udine, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Young, A.C., Shevchenko, Y.,
Sanchez, A., Whiting, M., Madan, A., Touchman, J.W., Green, E.D.,
Bouffard, G.G., Blakesley, R.W., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalilus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Mammalian Gene Collection Program Team
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences

CONSRM
TITLE

Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 16899-16903 (2002)

JOURNAL
PUBMED

NTH MGC Project
Direct Submission
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

REFERENCE
AUTHORS

Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

CONSRM
TITLE

Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 16899-16903 (2002)

JOURNAL
PUBMED

NTH MGC Project
Direct Submission
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

CONSRM
TITLE

Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 16899-16903 (2002)

JOURNAL
PUBMED

NTH MGC Project
Direct Submission
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

CONSRM
TITLE

Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 16899-16903 (2002)

JOURNAL
PUBMED

NTH MGC Project
Direct Submission
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

CONSRM
TITLE

Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 16899-16903 (2002)

JOURNAL
PUBMED

NTH MGC Project
Direct Submission
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

CONSRM
TITLE

Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 16899-16903 (2002)

JOURNAL
PUBMED

NTH MGC Project
Direct Submission
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

CONSRM
TITLE

Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 16899-16903 (2002)

JOURNAL
PUBMED

NTH MGC Project
Direct Submission
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

CONSRM
TITLE

Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 16899-16903 (2002)

JOURNAL
PUBMED

NTH MGC Project
Direct Submission
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

CONSRM
TITLE

Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 16899-16903 (2002)

JOURNAL
PUBMED

NTH MGC Project
Direct Submission
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

CONSRM
TITLE

Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 16899-16903 (2002)

JOURNAL
PUBMED

NTH MGC Project
Direct Submission
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

CONSRM
TITLE

Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 16899-16903 (2002)

JOURNAL
PUBMED

NTH MGC Project
Direct Submission
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

THIS PAGE LEFT BLANK

sb

protein"
 /protein_id="AAH21749.1"
 /db_xref="GI:18256806"
 /db_xref="GeneID:99167"
 /db_xref="MGI:2139150"
 /translation="MGDMVTVDVPLCTENKILSOYSETKMSPLSYSOQVLCSSVP
 LSKNVHGVFGVCTGENIEQISLKLSDMDHLOSCYAKLKEOLETSREMLGQERDR
 OLOCKNLSLHQLLKNEDVQKQNIIASRATQYNHVDKREINLKLKRLHQLVNN
 KXDKNDVINTYGRADGRKRSWRTDKTEARNEDEMYKILLNDYEQRIILMENAE
 LKKVLOQMKEMISLSPPQKKPRERAEDGTVAISDIEDDSGEISLSDSVGLSCDT
 VREGLTNSIRKQWRLKSHVEKLDNQASKVHSEGLNEEDVISRDHDEOETKLELEIE
 RCKEMIKAQOOLQOOLATTCDDDTSLRDYLLBEKRLKEWTLFKPKKNFERE
 RRSFTEAAIRGLERKAFEBERASVKKQFLNMTNPHQNSENVKLFSAFSGSDPDN
 LIVHSRQKQLSHVANGVACYSKLTKSPASPSISDFRQTHSCVSEHSSISVLNIT
 FEESKPSEVARESTDQKWSVQSRSPSSREGCYSGSSAFRSAHGRDRLP"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 3425
 Score: 3165.00 Matches: 615
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 6 Gaps: 0

US-10-644-084-2 (1-615) x BC021749 (1-3425)

QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
 DB 240 ATGGGAGATTGGATGCTGTCAGATCCAGTCTCTGTGTACAGAAACAAATCTCTCT 299
 QY 21 GlnTyThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
 DB 300 CAATATACCTCAGAAACAAAGATGTCTCCGTCAGTTGTACTCCAGCAAGTGTCTGTC 359
 QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
 DB 360 TCTTCAGTACCTTATCCAAAACGTCGATGGTGTCTTCGGTGTCTTCGACAGAGAG 419
 QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
 DB 420 AACATTGAACAAAGTATTTCTATCTTGTATCAGGAGCTGACACCTTCGGGTTCCTTCC 479
 QY 81 LeuTyrGluGlnSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
 DB 480 TTGTATGAAGATCCAAAGTAAAGAGGCAAGAGAGAAATTAATATAGTCGCTGTCTG 539
 QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
 DB 540 AACTGTATGAACGAGCTGCTGCTGCTCAGCGGAAGAACTTGTGCGCCAGAGAGCGTG 599
 QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
 DB 600 GAGACACAGAACTTGAAGCTGGGACGTGACATGGACCACTGACAGAGCTGTACGCCAA 559
 QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
 DB 660 CTTAAGGAGCAGTTGGAACGTCAGCGCGGAGATGATCGGGCTTCAAGAGAGAGACAG 719
 QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180
 DB 720 CAGCTCAGTGCAGAACAGAGAGTTGTCATCAGCTCTCTGAAGAAATGAGAGTGA 779
 QY 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
 DB 780 CAAAATTTACAAATATATATAGCCAGCGGGCTACTCAGTATATCATGATGTGAAGAG 839
 QY 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
 DB 840 AAGGAGCGTGAATATATAAGCTTAAAGGAGCGGCTGCATCAGCTCGTTATGAACAGAG 899
 QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240

DB 900 GATAAAAAATAGCCATGATGATTTTAAATATATGTGGTTCGAGCTGATGCAACACGAGC 959
 QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
 DB 960 TCATGAGGAGCTGACAAAACAGAGCCAGGATGAAGATGAGATGTACAAAATCTGTGTTG 1019
 QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
 DB 1020 AATGATTATGATGATCCAGGAGAGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1079
 QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
 DB 1080 CTCACAGATGAAGAAGAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1139
 QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320
 DB 1140 GAAAGAGCAGAGGAGCGCACAGGCACTGTTGCTATCTCCGATATAGAGATGATCTCTGG 1199
 QY 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
 DB 1200 GAACTGAGCAGAGACAGCGTGTGGGCGCTTATGAGGAGGAGCTCTCTCTCAGCAAGACCAT 1259
 QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
 DB 1260 AACAGCATCAGGAACACAGTGGAGAAATTTGAAAAGTCATGTAGAAAACTCGATAACCAA 1319
 QY 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
 DB 1320 GCTTCGAGGTACACTCAGAGGCGCTTATGAGGAGGAGCTCTCTCTCAGCAAGACCAT 1379
 QY 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400
 DB 1380 GAGCAAGAGACTGAGAAACTGAGAGTGGAGATTGAGCGGTGTAAGAGATGATCAAGGCT 1439
 QY 401 GlnGlnGlnLeuGlnGlnGlnLeuAlaThrCysAspAspAspThrThrSerLeu 420
 DB 1440 CAGCAGAGCTCTTACAGCAGCAGCTGGCCACCACTGTGATGATGATGATGATGATGATGAT 1499
 QY 421 LeuArgAspCysTyrLeuLeuGluGluArgGluLeuLysGluGluGluTrpThrLeuPhe 440
 DB 1500 TTGCGAGAGCTGTACTTCTGCGAAGAAAGAAAGCGCTTAAAGAGAGTGGACCTTTT 1559
 QY 441 LysGluGlnLysLysAspPheGluArgGluArgSerPheThrGluAlaIleArg 460
 DB 1560 AAAGAGCAAAAAAGAAATTTTGAAGAGAAAGCGGAGCTTTACAGAGCTGCCATTCCA 1619
 QY 461 LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnPhe 480
 DB 1620 TTGGGGTTGGAGAGAAAGCGCTTTGAAGAGAGCGGAGCGCTGGGTAAAGCAGCAGTTT 1679
 QY 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
 DB 1680 TTAAACATGACGAACTTTGACCAACAGAACTCAGAAAATGTGAAAATCTTTTCAAGTCTTTC 1739
 QY 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
 DB 1740 TCAGGAAGTTCTGATCCAGACAACTTATAGTCCACTCAGCGCCACCGCAAGAAAGCTTA 1799
 QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
 DB 1800 CACAGTGTGGCTTAAATGGGGTGCAGCTTGCACATCAAAACCTGACTTAAATCTCTCTCTGCC 1859
 QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
 DB 1860 TCACCTTCTACTTCAGACTTTGCGCAGACACATTCATGTGTGTCTGAACACAGTTCCATC 1919
 QY 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580
 DB 1920 AGTGTGCTGAATATAAATCTCTGAAAGAAAGTAAACCAAGTGGTGTGCAAGAGAAAGCAGC 1979
 QY 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCys 600

THIS PAGE LEFT BLANK

Db 1980 GATCAGAGTGGAGCGTGCAGTCGAGGCCAGCTCGCGGAGGGGTCTACAGCGGATGC 2039
 Qy 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
 Db 2040 TCCTCGGCTTTCAGGAGCGCTCAGCGGAGCCAGCAGATGATGATCTACCT 2084

BC031527 3410 bp mRNA linear ROD 29-JUN-2004
 Mus musculus synovial sarcoma, X breakpoint 2 interacting protein,
 mRNA (cDNA clone MGC:28268 IMAGE:4008624), complete cds.

ACCESSION BC031527
 VERSION BC031527.1 GI:21594536
 KEYWORDS MGC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 3410)
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schaefer, T.E., Brownstein, M.J., Udgin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE human and mouse cDNA sequences
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 PUBLISHED 12477932
 REFERENCE 2 (bases 1 to 3410)
 Strausberg, R.
 Direct Submission
 Submitted (06-JUN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Speed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 36 Row: k Column: 10
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 20270272.
 Location/Qualifiers
 1..3410
 /organism="Mus musculus"
 /mol_type="mRNA"

FEATURES
 source

gene

CDS

/strain="Czech II"
 /db_xref="taxon:10090"
 /clone="MGC:28268 IMAGE:4008624"
 /tissue_type="Mammary tumor metastasized to lung."
 MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR
 enhancer."
 /clone_lib="NCI CGAP_Lu30"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 1..3410
 /gene="Ssx2ip"
 /note="synonym: Adip"
 /db_xref="GeneID:99167"
 /db_xref="MGI:2139150"
 248..2095
 /gene="Ssx2ip"
 /codon_start=1
 /product="synovial sarcoma, X breakpoint 2 interacting
 protein"
 /protein_id="AAH31527.1"
 /db_xref="GI:21594537"
 /db_xref="MGI:2139150"
 /translation="MGDMVTVDVPLCTENKLSQYTSKMSPSLSYQVLCSSVP
 LSRNVGVGFCFTEQESYSLDQLDTFTGFPSPSLYEESKSEKRLNIVAVLNC
 MNELVLRQKLLAQESVETQNLKLGSDMDHLSQAKLKEQLTSRREMIGLQSRDR
 QLOCKNRSLLHOLKNEDEKQNI IASRATQYNDHVKRERKYLKRLHQLVMN
 KDKNIANDLVNTYGRADKGGKSWRTDKTEARNEDEMYKILLNDYKQILMENAR
 LKVVQKKKEMHISLLSPQKKPRERADGTCTVAISDIEDDSGLSRDSVWGLSCDT
 VREQNTSIRKQWRILKSHVEKLDNQASKVSEGLNEEDVSRQDHEQTEKLEIE
 RCKEMIKAQQLLOLQATTCDDTTSLRDCYLLEERKEWTLPKQKKNFPERE
 RRSFTAAIRLGLERKAPFEERASVWVKQOFLNMTFNDHONSENKLFSPSGSSDPDN
 LIVHSRPROKLLHVSANGVPACTSKLTSLPASPTSDPQTHSCVSEHSSISVLTIT
 PRESKPEVARESTQKWSVQSPRSRSGCYSGCSAFRSAHGRDRDLP"

ORIGIN

Alignment Scores: 0 Length: 3410
 Pred. No.: 0 Matches: 614
 Score: 3162.00 Conservative: 1
 Percent Similarity: 100.0% Mismatches: 0
 Best Local Similarity: 99.8% Indels: 0
 Query Match: 99.9% Gaps: 0
 DB: 6
 US-10-644-084-2 (1-615) x BC031527 (1-3410)
 Qy 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAenLeuSer 20
 Db 248 ATGGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTACAGAAACAAAAATCTCTCT 307
 Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
 Db 308 CAATATACCTCAGAACCAAGAGTGTCTCCCTCCAGTTTGTACTCCACGCAAGTTCTGTGC 367
 Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
 Db 368 TCTTCAGTACCTTTATCCAGAAATGTGATGTTTTCGGTGTCTTCTTCACAGAGAG 427
 Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
 Db 428 AACATTGAACCAAGATTTCTCTATCTGTATCAGAGCTGACCACTTCGGGTTCCTTCC 487
 Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
 Db 488 TTGTATGAAGAAATCCAAAGTAAAGAGGCAAGAGAGAAATTAATATATAGTCTCTTCTG 547
 Qy 101 AsnCysMetAsnGluLeuValLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
 Db 548 AACTGTATGAACGAGCTGCTGTGTTTCAGCGGAAGAACCTGCTGCTCCACAGAGAGCTG 607
 Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
 Db 608 GAGACACAGAACTTGAAGCTGGGCGAGTGACATGGACCACTTCAGAGCTGCTACGCCAAA 667

THIS PAGE LEFT BLANK

6a

OS Homo sapiens.
XX EP1308459-A2.
XX 07-MAY-2003.
XX 28-MAR-2002; 2002EP-00007401.
XX 05-NOV-2001; 2001JP-00379298.
XX 25-JAN-2002; 2002US-00350978.
XX (HELI-) HELIX RES INST.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX WPI; 2003-450961/43.
XX N-PSDB; ADB63343.
XX New polynucleotides and polypeptides, useful for developing a diagnostic
XX marker or medicines for regulation of their expression and activity, or
XX as targets of gene therapy.
XX Claim 1; Page; 222pp; English.
XX The invention discloses a polynucleotide comprising a sequence selected
XX from 1970 fully defined nucleotide sequences which encode novel
XX polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
XX or its partial peptide, an antibody binding to the polypeptide or peptide
XX of the polynucleotide, immunologically assaying the polypeptide or
XX peptide of the polynucleotide by contacting the polypeptide or peptide
XX with the antibody of the encoded protein, and observing the binding
XX between the two, a transformant carrying the polynucleotide in an
XX expressible manner and an antisense polynucleotide. The oligonucleotide
XX is useful as a primer for synthesizing the polynucleotide, or as a probe
XX for detecting the polynucleotide. The polynucleotides and encoded
XX proteins are useful as pharmaceutical agents and many disease-related
XX genes may be included in them, for developing a diagnostic marker or
XX medicines for regulation of their expression and activity, or as targets
XX of gene therapy. The genes are involved in tissue and/or cell
XX regeneration. Membrane proteins, signal transduction-related proteins,
XX transcription-related proteins, disease-related proteins and genes
XX encoding them can be used as indicators for diseases (e.g. osteoporosis,
XX neurological diseases, cancer, tumours. The cDNA may be used to regulate
XX the activity or expression of the encoded protein to treat diseases. The
XX sequence presented is a protein of the invention. Note: Some of the
XX sequence data for this patent is not represented in the printed
XX specification, but is based on sequence information supplied by the
XX European Patent Office.
XX Sequence 504 AA;
SQ Query Match 68.0%; Score 2152.5; DB 7; Length 504;
Best Local Similarity 90.3%; Pred. No. 1.6e-143;
Matches 419; Conservative 20; Mismatches 20; Indels 5; Gaps 2;
QY 1 MGDWMTVPDLCTENKNI:SOYTSKMPSSLYSQVLCSSVPLSKNVHGVGVECTGE 60
DB 1 MGDWMTVPDP----ESKTSQYTSKMPSSLYSQVLCSSVPLSKNVHGVGVECTGE 56
QY 61 NIEQSIYLDQELTTGFFPSLYEESKAKREINIVAVLNCNELLVLQRNLLAQESV 120
DB 57 NIEQSIYLDQELTTGFFPSLYEESKAKREINIVAVLNCNELLVLQRNLLAQENV 116
QY 121 ETQNLKSGDMHLOSCYAKLKEQLETSREMIQGLQERDRQLQCKNRLHQLLNKEKDE 180
DB 117 ETQNLKSGDMHLOSCYAKLKEQLETSREMIQGLQERDRQLQCKNRLHQLLNKEKDE 176
QY 181 OKLQNIASRATQYNDHVRKEREYNKLERHQLVWKKDKNIAMDVLNYVGRADKRG 240
DB 177 OKLQNIASRATQYNDHVRKEREYNKLERHQLVWKKDKNIAMDVLNYVGRADKRG 236

QY 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAEKLVQOMKEMISLLSPQKKPR 300
DB 237 SWRTGKTEARNEDEMYKILLNDYEQKQILMENAEKLVQOMKEMISLLSPQKKPR 296
QY 301 ERAEDGTGVAISDIEDSGELSRDSVWGLSCDTVRBQLTNSIRKQWRILKSHVEKLDNQ 360
DB 297 ERVDDSTGTV-ISDVEEDAGELSRDSVWGLSCDTVRBQLTNSIRKQWRILKSHVEKLDNQ 355
QY 361 ASKVHSEGLNEEDVTSRODHEQTEKLELEIERCKEMIKAQOQLLOQLATTCDDDTTSL 420
DB 356 VSKVHLEGFNEDVTSRODHEQTEKLELEIOCKEMIKAQOQLLOQLATTCDDDTTSL 415
QY 421 LRDCCVLLERKLEKBEWTLFKQKKNFERERSFTEAAILGLIE 464
DB 416 LRDCCVLLERKLEKBEWTLFKQKKNFERERSFTEAAILGLIE 459
RESULT 5
AAB93250
XX AAB93250 standard; protein; 417 AA.
XX AC AAB93250;
XX DT 26-JUN-2001 (first entry)
XX DE Human protein sequence SSQ ID NO:12265.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-00116126.
XX PR 29-JUL-1999; 99JP-00248036.
XX PR 27-AUG-1999; 99JP-00300253.
XX PR 11-JAN-2000; 2000JP-00118776.
XX PR 02-MAY-2000; 2000JP-00183767.
XX PR 09-JUN-2000; 2000JP-00241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX PS Claim 8; SSQ ID NO 12265; 2537pp + Sequence Listing; English.
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the

THIS PAGE LEFT BLANK

CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
xx
SQ Sequence 417 AA;

Query Match	58.6%	Score 1855.5	DB 4	Length 417
Best Local Similarity	85.6%	Pred. No. 1.3e-122		
Matches 358	Conservative 27	Mismatches 32	Indels 1	Gaps 1
Qy	198	VKKEREYNKLERLHOLVMNKDKNTAMDVLNVYGRADGKRGSWRTDKTEARNEDEMYK	257	
Db	1	MKKEREYNKLERLHOLVMNKDKNTAMDVLNVYGRADGKRGSWRTDKTEARNEDEMYK	60	
Qy	258	ILLNDYEQKQILMENAELKKVLLQMKKEMISLLSPQKKPRERAEDGTGTVASIDIED	317	
Db	61	ILLNDYEQKQILMENAELKKVLLQMKKEMISLLSPQKKPRERVDSTGTV-ISDVEE	119	
Qy	318	DSGELSRSVWGSLCSDTVREQLTNSIRKQWRILKSHVEKLDNQAKVHSGLENEEDVISR	377	
Db	120	DAEELSRESHWDJSCETVREQLTNSIRKQWRILKSHVEKLDNQAKVHLEGFENDEVISR	179	
Qy	378	QDHEQETKLELEIERCEKMIKAOQQLLOQLATTCDDMTTSLLRDCYLLKEERLKEEW	437	
Db	180	QDHEQETKLELEIQOQCEMIKTOQQLLOQLATAYDDTTSLLRDCYLLKEERLKEEW	239	
Qy	438	TLPEKQKQKFNERRRSFTEAAIRLGLERKAFEBERASVWQOFLNMTNPFQHNSENVKLF	497	
Db	240	SLFPEKQKQKFNERRRSFTEAAIRLGLERMAFEERASWLKQOFLNMTTTFQHNSENVKLF	259	
Qy	498	SAPSGSSDDPNLIVHSPROKHLHSAVANGVPACTSKTLKSLPASPSTSDPRQTHSCVSEH	557	
Db	300	SAPSGSSDDNLLVHSPQPOKPPHVSNGSPVCMKSLTKSLPASPSTSDFCQTRCLISEH	359	
Qy	558	SSISVLNITPEESKPESEVARESTQKWSVOSRPSREGCVSGGSSAFRSHAGDRDDLP	615	
Db	360	SSINVLNITAEETKPNQGVGECTNQKWSVASRPGSQEGCVSGGSLSYTNSHVRKDDLP	417	

RESULT 6
AAG03000
ID AAG03000 standard; protein; 137 AA.

XX	06-OCT-2000 (first entry)
DT	Human secreted protein, SEQ ID NO: 7081.
XX	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation; gene therapy; chromosome mapping.
DE	
XX	
XX	
KW	
KW	

at	OS	Homo sapiens.
XX	XX	
PN	PN	EP1033401-A2.
XX	XX	
PD	PD	06-SEP-2000.
XX	XX	
PF	PF	21-FEB-2000; 2000EP-00200610.
XX	XX	
PR	PR	26-FEB-1999; 99US-0123487P.
XX	XX	
PA	PA	(GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
DR N-PSDB; AAC03006.
DR

DR WPI; 2000-500381/45.
DR N-PSDB; AAC03006.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
 PS
 PS Claim 13; SEQ ID NO 7081; 71pp + Sequence Listing; English.
 XX
 CC The present sequence is a polypeptide encoded by one of a large number of
 CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
 CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
 CC tissues. EST sequences usually correspond mainly to the 3' untranslated
 CC region (UTR) of the mRNA because they are often obtained from oligo-dT
 CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
 CC sequences derived from the 5' ends of mRNAs and even in those cases where
 CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
 CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
 CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
 CC are also used in diagnostic, forensic, gene therapy and chromosome
 CC mapping procedures. They are used to obtain upstream regulatory sequences
 CC and to design expression and secretion vectors
 XX
 XX Sequence 137 AA:
 SQ

```

Query Match      . 19.4%; Score 612.5; DB 3; Length 137;
Best Local Similarity 87.0%; Pred. NO. 2.4e-35;
Matches 120; Conservative 10; Mismatches 7; Indels 1; Gaps 1;

Qy 272 MENAEELKVLQOMKKEMISLLSPQKKPRERADGTGTVAISDIEDSDGSELSDSWGLS 331
      |||||
Db 1 MENAEELKVLQOMKKEMISLLSPQKKPRERVDSDGTGTV-ISDVEEDAGELSGRESMWDL 59
      |||||

Qy 332 CDTVREQLTNSIRKQWRILKSHVEKLDNQASKVHSEGLNEEDVISRDHQEQTEKLELBI 391
      |||||
Db 60 CETVREQLTNSIRKQWRILKSHVEKLDNQOVSKVHLEGFNDVEDVISRDHQEQTEKLELBI 119
      |||||

Qy 392 ERCKEMIKAQOOLLOOOL 409
      ::|||
Db 120 QOCKEMIKTQOOLLOOOL 137
      |||||

```

RESULT 7
ADX91485

ID ADX91485 standard; protein; 443 AA.

AC ADX91485;

DT 21-APR-2005 (first entry)

DE Plant full length insert polypeptide seqid 54149.

plant protectant; plant growth regulant; gene therapy; plant;
recombinant DNA construct; physical array; plant breeding marker;
cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
extreme osmotic condition; pathogen tolerance; pest tolerance;
growth rate; cell cycle pathway; disease resistance;
galactanmannan production; lignin production; plant growth regulator;
yield; plant growth; plant development; seed oil; protein yield;
protein content.

Unidentified.

AA
PN
US2004034888-A1.XX
PD
19-FEB-2004.

28-APR-2003; 2003US-00425114.

PR 06-MAY-1999: 99US-00304517.

FR 03-NOV-2001; 2001US-009856/8.
XX

PA (LIU/) LIU J.
PA (ZHOU/) ZHOU Y.

PA (ZHOU/) ZHOU I.
PA (KOVA/) KOVALIC D K.

THIS PAGE LEFT BLANK

1007	GAACGTG	1063	TTAAATCTATTAAAGGATATAAAAGCTTTGGATATGTTATTTTAGTAACAGAGCATC	2143
1062	GGGCGCTTCTGTCAGTGTGAGAGAGCAGCTGTACAAACAGCATCAGGAACAGTGA	1121	TGGCTCTGTGAGCTGAGGATGTCAT - GCGTTTGGATGGAACAGAGCGCTGGATGA	2178
1064	GGGACCTTCTGTCAGTGTGAGAGAGCAGCTGTACAAACAGCATCAGGAACAGTGA	1123	TGGCTCTGTGAGCTGAGGATGTCATAGATGTTTGGATGGAACAGAGCGCTGGATGA	2203
1122	GAATTTTGAAGTCTATGTAGAAAACTCGATACCAAGCTTCGAGGTTCTGAGGTTACCTCAGAG	1181	-TTGCTGCTGCCAGGTACCGAGAGAGCAGCTTTTAGGAGCTGCTTCTGTAAACATTAAT	2237
1124	GAATTTTGAAGTCTATGTAGAAAACTCGATACCAAGCTTCGAGGTTCTGAGGTTACCTCAGAG	1183	TTTCTCTTATAGGTAATTAATAATAGCAGCTTTTAGGAACTGATATTGTAATGTTTAA	2263
1182	GCCTTAATGAGGAGGCTCATCTCAGCAAGACCATGAGCAGAGAGCTGAGAACTGG	1241	ATTGCTCCCAAGTGTGTTGGCAATTTGGAAGTGTAGCTTTTACTTTGAATGTATCTGTAGA	2297
1184	GTTTTAATGATGAAGTGTAACTCAGCAGACCAAGATGAAGAACTGGAAGAACTCG	1243	TTTGTCTCAAAATATAGTTGGCAATTTGGAAGTGTAGCTTTTACTTTGAATGTATCTGTAGA	2323
1242	AGCTGGAGTGTGCGGTGTAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGC	1301	TTTTTAAACAAAGCAGGTTCTATATTATTATGTTTATGTTTGGGATTTGGGATTTACCTCTT	2357
1244	AGTTAGAAATTCAGCAGTGTAAAGAAATGATTAATACTCAGCAACAGCTTTTACAGCAGC	1303	TTTTTAAACAAAGCAGGTTCTATATTATTATGTTTATGTTTGGGATTTGGGATTTACCTCTT	2382
1302	AGCTGGCCACACGCTGTGATGATGACACACCTCCTGTTGCGAGACTGTTTACTTGTCTG	1361	-----GTTGCTGTACATAATAATATATGATGTTTGTATACATAGATACACATGATTTG	2442
1304	AGCTCGCTATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1363	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	2456
1362	AAGAAAAGGAGCGCTTAAAGAGAGTGGACCTTTTAAAGAGCAAAAGAAATTTG	1421	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	2500
1364	AAGAAAAGGAGCGCTTAAAGAGAGTGGACCTTTTAAAGAGCAAAAGAAATTTG	1423	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	2514
1422	AGAGAGAGGCGGAGCTTACAGAGCTGCGATTCGATTTGGGTTGGAGAGAAAGCGGT	1481	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	2560
1424	AGAGGAGAGAGCGAGCTTACAGAGCTGCGATTCGATTTGGGTTGGAGAGAAAGCGGT	1483	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	2573
1482	TTGAGAGAGAGCGGAGCTTAAAGAGAGTGGACCTTTTAAAGAGCAAAAGAAATTTG	1541	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	2619
1484	TTGAGAGAGAGCGGAGCTTAAAGAGAGTGGACCTTTTAAAGAGCAAAAGAAATTTG	1543	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	2629
1542	ACCAGAGCTCAGAAAATGTGAACCTTTTCAAGTGCCTTCTCAGGAAGTTCTGATCCAGACA	1601	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	2679
1544	ACCAGAGCTCAGAAAATGTGAACCTTTTCAAGTGCCTTCTCAGGAAGTTCTGATGGGACA	1603	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	2688
1602	ATCTTATAGTCACTCAGCGGACCGCAAGAGCTACACAGTGTGGCTTAATGGGGTGC	1661	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	2739
1604	ATCTTATAGTCACTCAGCGGACCGCAAGAGCTACACAGTGTGGCTTAATGGGGTGC	1663	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
1662	CAGCTTGACATCAAACTGACTTAAATCTCTTCTGCGCTCAGCTTCTCAGAGTTTC	1721	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
1664	CAGTTTGCATGCTAACTTACTTAATCTCTTCTGCTTCACTTCCACTTCCAGACTTTT	1723	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
1722	GCCAGACATTCATGCTGTGTAACACAGTTCATCAGTGTGCTGATTAATACCTCTG	1781	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
1724	GCCAGACATTCATGCTGTGTAACACAGTTCATCAGTGTGCTGATTAATACCTCTG	1783	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
1782	AGAAAAGTAAACCAAGTGTGTAACACAGTTCATCAGTGTGCTGATTAATACCTCTG	1841	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
1784	AGAAAAGTAAACCAAGTGTGTAACACAGTTCATCAGTGTGCTGATTAATACCTCTG	1843	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
1842	CGAGGCCAGCTCGCGGAGGGTGTCTACAGCGATGCTCTCGGCTTCCAGAGGCTC	1901	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
1844	CAAGACTGGATCACAGGAGGTTGCTTATAGTGTGCTCTCTTCCAGCTACACAAATCTC	1903	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
1902	ACGGGACCGAGATGATTAACCTTAAATGTGCGGGTGCAGTGTCTTCCAGATGTGCG	1961	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
1904	ATGTAGAAAAGAGATGATTAACCTTAAATGTGCGGGTGCAGTGTCTTCCAGATGTGCG	1963	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
1962	CTAGAGG-----AGTTGACAGAGGTTGTAGCATTAAGTCAAGTCAAGTCTGCTC	2005	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
1964	CATCAAGTTTCACTAAGTTGAAACAGGTTGTGTCTAATAGTCAAGTCTAATATAA	2023	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
2006	CTTAAGATGCTCAGAGTGTGTTTGTGAGTCTGCTGCTTCCCGGAGAGCTGAATG	2065	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
2024	CTTAAGATGCTGAGTGTGTTTGTGAGTCTGCTGCTTCCCGGAGAGCTGAATG	2083	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	
2066	CT-AAGCTACTTAAAGGATGCAAGCTTTGG-----TTGTGTGTAGTAAACAGAGCCCC	2120	TTAAGGCGCTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAAAGGTTTAA	

RESULT 4

AAH14625
ID AAH14625 standard; cDNA; 2716 BP.
XX AC AAH14625;
XX XX 26-JUN-2001 (first entry)
XX XX Human cDNA sequence SEQ ID NO:12264.
XX XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX XX Homo sapiens.
XX XX EP1074617-A2.
XX XX 07-FEB-2001.
XX XX 28-JUL-2000; 2000EP-00116126.
XX XX 29-JUL-1999; 99JP-00248036.
XX XX 27-AUG-1999; 99JP-00300253.
XX XX 11-JAN-2000; 2000JP-00118776.
XX XX 02-MAY-2000; 2000JP-00183767.
XX XX 09-JUN-2000; 2000JP-00241899.

THIS PAGE LEFT BLANK

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 12264; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2716 BP; 935 A; 405 C; 524 G; 852 T; 0 U; 0 Other;

SV Match	46.8%;	Score 1260.4;	DB 4;	Length 2716;
- Local Similarity	79.7%;	Pred. No. 0;		
Matches 1716; Conservative	0;	Mismatches 376;	Indels 60;	

[illegible]

THIS PAGE LEFT BLANK

QY 2133 TGCAGGAATGCAT-GGCGTTTGGATGGAACAGAGCGCTGGAATGA--TTGCCTGCCA 2189
 Db 1558 TAAAGGAATGTATAGATGTTTGGATGGAACAAACAGCACTAGACTGAGTTTCTCTTATA 1617
 QY 2190 GGTACCCAGAGAGACACTTTTGGGACTGGTTCCTGTAACATTAATAATTCGTCCCAAG 2249
 Db 1618 GGTATTAATAATAGCACTTTTGGAACTGATTAATGTTAAATTTGTCCTCAA 1677
 QY 2250 TGTGTTGGCAATGGAAGTGAAGCTTTTACCTTGAATGTATCTAGATTTTAAACAAAG 2309
 Db 1678 TATAGTTGGCATGGAAGTTTACCTTTTACTTGAATGTATCTAGATTTTAAACAAAG 1737
 QY 2310 CAGGTTCTATATTTATTTATGTTTATGTTTGGGATTAACCTCTTTTATATGTTT-- 2367
 Db 1738 CGAGTTCTATATTTATTTATGTTTATGTTG-GTTTTGAATTAACCTCTTTTATATGTTTAA 1796
 QY 2368 -----GTGTCGTACATAAATAATACATGATGTTTAAAGAGGCTTT 2408
 Db 1797 ATAAAGTGAATTTATGTTTATGTTTGTACATAGATACATGATTTAAAGAGGCTTT 1856
 QY 2409 AAGGTTTAAACCTTCAACCATGCTTGAAGTATAGCATTTTCAAGCAATTAATAATGTTT 2468
 Db 1857 AAGATTTAAAGTTTCA-CAACCATAGTATAGTATTTTATGCTCC-AGTAAATTTT 1914
 QY 2469 CAGTGGCATGTTTGTACAGAGG--TTAGGACCACTGCCATCAGATTAAGACTTTAT 2526
 Db 1915 TAGTGGTATTTCTGTTTACAGATGTTTATGAGGACCATGATGATTTTAAAG-ATTCT 1973
 QY 2527 TTTTAAAGCATCTGGGCAATAAATAATCAAGCCCTTTCAT-AAAGTGAAGTTTC---AGAT 2582
 Db 1974 CTTTTAAATACATCTGGGCAATAAATAATGAAAGGTATTTCCATGAAGCTGAGTTCTTTAGAT 2033
 QY 2583 AACTAGAACTACTAACCTTACATTTTGGAGATTTTAAAGCAATTTTAAATTTTATA 2641
 Db 2034 AATCAACACTACTAACATTTTGGAGATTTTATGATTTTATGATTTTATTTTGTGA 2093
 QY 2642 TATGT-GAATGTTTAAATTTTCAAGAGGAATATTTGATTTGAGTAAATGGG 2692
 Db 2094 TATGTAGAAATTTATTAATTTTAAAGGACTATTGATGATAGAAATAGGG 2145

RESULT 5

ADB63343
 ID ADB63343 standard; cDNA; 2272 BP.

AC ADB63343;

04-DEC-2003 (first entry)

Human cDNA encoding clone TESTI20071130.

Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
 tissue regeneration; cell regeneration; membrane protein;
 signal transduction-related protein; transcription-related protein;
 osteoporosis; neurological disease; cancer; tumour.

Homo sapiens.

Key Location/Qualifiers
 CDS 224..1738

/*tag= a
 /product= "Clone TESTI20071130 protein"

EP1308459-A2.

07-MAY-2003.

28-MAR-2002; 2002EP-00007401.

05-NOV-2001; 2001JP-00379298.

25-JAN-2002; 2002US-00350978.

(HELI-) HELIX RES INST.
 (REAS-) RES ASSOC BIOTECHNOLOGY.

Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 WPI: 2003-450961/43.
 P-PSDB; ADB65313.

New polynucleotides and polypeptides, useful for developing a diagnostic
 marker or medicines for regulation of their expression and activity, or
 as targets of gene therapy.

Claim 1; Page; 222pp; English.

The invention discloses a polynucleotide comprising a sequence selected
 from 1970 fully defined nucleotide sequences which encode novel
 polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 or its partial peptide, an antibody binding to the polypeptide or peptide
 of the polynucleotide, immunologically assaying the polypeptide or
 peptide of the polynucleotide by contacting the polypeptide or peptide
 with the antibody of the encoded protein, and observing the binding
 between the two, a transformant carrying the polynucleotide in an
 expressible manner and an antisense polynucleotide. The oligonucleotide
 is useful as a primer for synthesizing the polynucleotide, or as a probe
 for detecting the polynucleotide. The polynucleotides and encoded
 proteins are useful as pharmaceutical agents and many disease-related
 medicines for regulation of their expression and activity, or as targets
 of gene therapy. The genes are involved in tissue and/or cell
 regeneration. Membrane proteins, signal transduction-related proteins,
 transcription-related proteins, disease-related proteins and genes
 encoding them can be used as indicators for diseases (e.g. osteoporosis,
 neurological diseases, cancer, tumours). The cDNA may be used to regulate
 the activity or expression of the encoded protein to treat diseases. The
 sequence presented is a cDNA of the invention. Note: Some of the sequence
 data for this patent is not represented in the printed specification, but
 is based on sequence information supplied by the European Patent Office.

Sequence 2272 BP; 783 A; 398 C; 502 G; 589 T; 0 U; 0 Other;

Query Match 46.8%; Score 1258.8; DB 10; Length 2272;

Best Local Similarity 77.2%; Pred. No. 0;

Matches 1651; Conservative 0; Mismatches 337; Indels 150; Gaps 4;

QY	1	CGTAGAGAGTGACAGAGAGCTGTTGTAAGCGTCGACAGCTAGCCGCTCCAGGAT	60
Db	145	CATAGAAGAGTGACAGCAGCTAGACTAATAATGTTTAACTGCTGAACCTAGTTCCTCAGGAT	204
QY	61	CCTGGCTCTGGAACCTTGCTATGGAGATTGCGATGCTGCACAGATCCAGTTCGTGTAC	120
Db	205	CCTGGCTCTGAGATTGCTATGGAGATTGCGATGCTGCATGCTTACAGATC-----C	252
QY	121	AGAAAAACAAATATCTCTCAATATACCTCAGAAACAAAGATGTCCTCGTCCAGTTGTA	180
Db	253	AGAAAGCAAACTATCTCTCAATATACCTCAGAAACAAAGATGTCCTCAACAGTTTATA	312
QY	181	CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTATCCAAAACGTGATGGTGTTCGG	240
Db	313	CTCAGCAAGTGTGTATGTTCTTCAATACCTTTATCGAAAATATGTGCACAGTTTTTCAG	372
QY	241	TGTTCTCTGCACAGGAGAGCAATTAAGCAAAAGTATTTCTTATCTGATCAGGAGCTGAC	300
Db	373	TGCTTCTCTGCACAGAGATTAATTTGAAACAGATATCTCATATCTGATCAGGATTTGAC	432
QY	301	CACCTTCGGGTTTCTTCTTGTATGAAGATCCAAAGTAAAGAGCAAGAGAGAAATT	360
Db	433	TACTTTTGGTTTCTTCTTATATGAAGATCCAAAGTAAAGAGCAAGAGAGAGTT	492
QY	361	AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTCTGTCGCTCAGCGGAGAACCT	420
Db	493	AAATATAGTAGCTGTACTAAATTTGTATGAATGAGCTGCTTGTGCTTTCAGCGGAGAACCT	552

THIS PAGE LEFT BLANK

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2006, 22:24:24 ; Search time 498 Seconds
(without alignments)
10114.508 Million cell updates/sec

Title: US-10-644-084-1

Perfect score: 2692

Sequence: 1 cgtaggagagcagcagcagc.....ttgatttgaggaatggggg 2692

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 0%

Maximum Match: 100%

Listing first 45 summaries

Database :

- 1: Issued Patents NA:*
- 2: /EMC_Celerra_SID33/ptodata/2/ina/1 COMB.seq:*
- 3: /EMC_Celerra_SID33/ptodata/2/ina/5 COMB.seq:*
- 4: /EMC_Celerra_SID33/ptodata/2/ina/6A COMB.seq:*
- 5: /EMC_Celerra_SID33/ptodata/2/ina/6B COMB.seq:*
- 6: /EMC_Celerra_SID33/ptodata/2/ina/7 COMB.seq:*
- 7: /EMC_Celerra_SID33/ptodata/2/ina/8 COMB.seq:*
- 8: /EMC_Celerra_SID33/ptodata/2/ina/9 COMB.seq:*
- 9: /EMC_Celerra_SID33/ptodata/2/ina/PP COMB.seq:*
- 10: /EMC_Celerra_SID33/ptodata/2/ina/RE COMB.seq:*
- 11: /EMC_Celerra_SID33/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1258.8	46.8	2272	3	US-10-104-047-1497
2	297.2	11.0	428	3	US-09-513-999C-3004
3	73.6	2.7	7218	2	US-08-232-463-14
4	58.2	2.2	198	3	US-09-513-999C-29101
5	55.8	2.1	2093	3	US-10-104-047-1666
6	50.8	1.9	1141	3	US-09-806-708B-22
7	48.8	1.8	5883	3	US-09-949-016-5001
8	48.8	1.8	28806	3	US-09-949-016-13217
9	46.6	1.7	1312	3	US-09-976-594-886
10	46	1.7	3489	2	US-08-728-323A-1
11	46	1.7	3489	3	US-09-298-568-1
12	46	1.7	3489	3	US-09-410-399-1
13	46	1.7	3489	3	US-09-894-273-1
14	46	1.7	32207	2	US-08-770-379-20
15	46	1.7	32207	3	US-08-757-669A-20
16	46	1.7	32207	3	US-09-230-371A-20
17	44.4	1.6	543	10	5273901-6
18	44.4	1.6	832	3	US-09-621-976-2813
19	43.6	1.6	16442	3	US-08-781-891-208
20	43.6	1.6	16442	3	US-09-618-166-208
21	43.4	1.6	2357	3	US-10-104-047-1780
22	43.4	1.6	15252	3	US-09-949-016-13584
23	43	1.6	3337	2	US-08-072-610-1

24	43	1.6	3337	2	US-08-719-822B-1	Sequence 1, Appli
25	43	1.6	3337	3	US-09-092-458-1	Sequence 1, Appli
26	43	1.6	3337	3	US-08-719-821C-1	Sequence 1, Appli
27	42.8	1.6	1995	2	US-08-425-069-3	Sequence 3, Appli
28	42.8	1.6	1995	2	US-08-317-844B-3	Sequence 3, Appli
29	42.4	1.6	1852	3	US-09-969-852-4	Sequence 4, Appli
30	42.2	1.6	150394	3	US-09-949-016-13042	Sequence 13042, A
31	42	1.6	1664976	3	US-08-916-421B-1	Sequence 1, Appli
32	42	1.6	1664976	3	US-09-692-570-1	Sequence 1, Appli
33	41.8	1.6	1745	3	US-09-949-016-4711	Sequence 4711, Ap
34	41.8	1.6	5746	3	US-09-949-016-16453	Sequence 16453, A
35	41.6	1.5	289	3	US-09-007-005-17	Sequence 17, Appl
36	41.6	1.5	289	3	US-09-244-796-17	Sequence 17, Appl
37	41.4	1.5	92227	3	US-09-949-016-11929	Sequence 11929, A
38	41.4	1.5	92227	3	US-09-949-016-15421	Sequence 15421, A
39	41.2	1.5	929	3	US-09-671-317-14	Sequence 14, Appl
40	41.2	1.5	1001	3	US-09-671-317-439	Sequence 439, App
41	41	1.5	2338	2	US-08-425-069-1	Sequence 1, Appli
42	41	1.5	2338	2	US-08-317-844B-1	Sequence 1, Appli
43	41	1.5	2532	3	US-09-328-352-1168	Sequence 1168, Ap
44	41	1.5	640681	3	US-09-790-988-1	Sequence 1, Appli
45	40.8	1.5	330	4	US-10-211-689-123	Sequence 123, App

ALIGNMENTS

RESULT 1
US-10-104-047-1497
; Sequence 1497, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cdna
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1497
; LENGTH: 2272
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1497

Query Match	46.8%	Score	1258.8	DB	3	Length	2272
Best Local Similarity	77.2%	Pred	No. 0				
Matches	1651	Conservative	0	Mismatches	337	Indels	150
Gaps	4						
Qy	1	CGTAGGAGGTGACAGGAGCTGTTGTAAGCGCTGCAGCAGCTGAGCGGCTCTCTCAGGTAT	60				
Db	145	CATAGAAGAGTGACAGCAGCTAGACTAAATGTTTAACTGCTGAAGTCTCTCAGGTAT	204				
Qy	61	CCTGGCTCTGGAAGTCTGCTATGGAGATTGAGTCTGTGACAGATTCAGTTCTGTGTAC	120				
Db	205	CCTGGCTCTGAGAGTCTGATGGAGATTGAGTCTGTGACATC-----C	252				
Qy	121	AGAAACAAAAATCTCTCAATATACCTCAGAACAAAGATGTCCTCGTCCAGTTGTA	180				
Db	253	AGAAAGCAAACTATCTCTCAATATACCTCAGAACAAAGATGTCCTCAATGTTTATA	312				
Qy	181	CTCCAGCAAGTCTGTGCTCTTCAGTACCTTTATCAAAAACTGATGCTGTTTCGG	240				
Db	313	CTCAGCAAGTCTGTGCTCTTCAGTACCTTTATCGAAAAATGTCAGAGTTTTCAG	372				
Qy	241	TGCTCTTTCAGCAGGAGAGCAATTGAACTATTTCTATCTTATCAGGAGCTGAC	300				
Db	373	TGCTCTTTCAGCAGGAGAGCAATTGAACTATTTCTATCTTATCAGGAGCTGAC	432				
Qy	301	CACCTTCGGGTTTCTCTTCTTGTATGAGGATCCAAAGTAAAGGCAAGAGAGATT	360				

'HIS PAGE LEFT BLANK'

433 TACTTTTGGTTTCTCTTCATTATATGAAGAATCAAGAGTAAAGAGACAAAGAGAGAGTT 492
QY
361 AATATAGTCTGCTGTTCTGAACTGTATGAACAGCTGCTCGTCTTCAGCGGAAGAACTT 420
Db
493 AATATAGTCTGCTGTTCTGAACTGTATGAACAGCTGCTCGTCTTCAGCGGAAGAACTT 552
QY
421 GCTGCGCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGGCGAGTGACATGGACCACT 480
Db
553 TCTAGCTCAGGAATATGTGGAGACACAGAACTTTGAAGCGGGAAGTATGACCACT 612
QY
481 CGAGAGTCTCTACGCCAAACTTAAGAGAGAGTGTGAAGAGCTCCAGCGGAGATGATCGG 540
Db
613 ACAGAGTCTCTACTCAAACTTAAGGAACAACCTGGAACCTCCAGGAGGGAATGATGG 672
QY
541 GCTTCAAGAGAGACAGCGAGCTGCAGTGCAGACAGAGCTTTGACATCAGCTCCTGAA 600
Db
673 GCTTCAGGAAGAGACAGAGCTGTAATGTAAGAACAGGAATTTGCAATCAGCTACTAAA 732
QY
601 GAATGAGAAAGATGAGGTAACAAAATTTACAAAATATCATAGCCAGCGGGCTACTCAGTA 660
Db
733 GAATGAGAAAGATGAGGTCGAAAATTTACAAAATATCATAGCTCAGCTACTCAGTA 792
QY
661 TAATCATGATGTAAGAGAGAGAGCGTGAATATTAATAGCTTAAAGAGCGGCTCATCA 720
Db
793 TAATCATGATGTAAGAGAGAGAGCGTGAATATTAATAGCTTAAAGAGCGGCTCATCA 852
QY
721 GCTCGTTATGAACAGAGAGTAAACATAGCCATGATGTTTAAATTTATGTTGGTCCG 780
Db
853 ACTTGTATGAACAGAGAGTAAAGAAATAGCTATGGAATTTTGAATTTATGTTGGGAG 912
QY
781 AGCTGATGCAACAGAGGCTCATGAGGAGTCAACAAAACAGAAAGCCAGGAATGAAGATGA 840
Db
913 AGCTGATGCAACAGAGGCTCATGAGGAGTCAACAAAACAGAAAGCCAGGAATGAAGATGA 972
QY
841 GATGATCAAAATTTCTGTTGAATGATTAAGTACCCAGAGAGAGAGTCTGATGAGAA 900
Db
973 AATGATCAAAATTTCTGTTGAATGATTAAGTACCCAGAGAGAGAGTCTGATGAGAA 1032
QY
901 CGCGAGCTGAGAGAGTCTTCCAGCAGATGAGAGAGAGTATCTCTCTCTCTCTCTCC 960
Db
1033 TGCAAGAACTTAAGAGAGTCTTCCAGCAGATGAGAGAGAGTATCTCTCTCTCTCTCTCC 1092
QY
961 TCAG 1020
Db
1093 CCAAG 1149
QY
1021 TATAGAGATGATCTGCGGAACTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Db
1150 TGTGAG 1209
QY
1081 TGTGAG 1140
Db
1210 TGTGAG 1269
QY
1141 AGAAAACTCGATAACAGCTTCAAGGTACACTCAGAGGCGCTTAAATGAGAGAGAGAG 1200
Db
1270 AGAAAACTCGATAACAGCTTCAAGGTACACTCAGAGGAGAGAGAGAGAGAGAGAGAG 1329
QY
1201 CATCTCAGCAG 1260
Db
1330 AATCTCAGCAG 1389
QY
1261 TAAAG 1320
Db
1390 TAAAG 1449
QY
1321 TGATGACACCACTCACTCTTCCGAGAGTCTTCTGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db
1450 TGATGATACCACTCACTCTTCCGAGAGTCTTCTGAGAGAGAGAGAGAGAGAGAGAG 1509
QY
1381 AGAAG 1440
Db
1510 AGAAG 1569

QY 1441 TACAGAGCTGCCATTCGATTGGGTTGGAG----- 1471
Db 1570 TACAGAGCCGCTATTCGCTGGGATTTGGAGATTTGGCTTTCTCTGCAAGTGTTTTACA 1629
QY 1472 ----- 1471
Db 1630 CAAGAAATCTCTTTTGGGTATCCATCATTCATCCCGTCAGTAGGTGTGGTGTGTCAGAAG 1689
QY 1472 -----AGAAAGCGTTTGAAGAGCGAG 1496.
Db 1690 ACAGTGCACCAAGACTATATTGCTGCTTTAAAAAAGAAAGGCAATTTGAAGAGAAAGAG 1749
QY 1497 CCAGCTGGTAAAGCAGCAGTGTAAACATGACGAACCTTTGACACCCAGAACTCAGAAA 1556
Db 1750 CCAGTGTGTTAAGCAGCAGTGTCTAAATATGACTACTTTGACACCCAGAACTCAGAAA 1809
QY 1557 ATGTGAACTTTTTCAGTGCCTTCTCAGAAAGTTCCTGATCCAGACAACTTTATAGTCCA 1616
Db 1810 ATGTGAACTTTTTCAGTGCCTTCTCAGAAAGTTCCTGATTTGGGACAACTTTATAGTCCA 1869
QY 1617 CAGGCGCAGGCAAGAGAGCTACACAGTGTGGCTAATGGGTGCTCCAGTTTGCATGCTA 1676
Db 1870 CGAGGCGAGCGCAAGAGAGAGCTCAGAGTGTGTCTAATGGGTCTCCAGTTTGCATGCTA 1929
QY 1677 AACTGACTTAAATCTTCTTCCCTCAGCTTCTACTTCTCAGACTTTTGGCCAGACACATTCAT 1736
Db 1930 AACTTACTTAAATCTTCTTCCCTCAGCTTCTACTTCTCAGACTTTTGGCCAGACACATTCAT 1989
QY 1737 GTGTGTCGAACACAGTTCCTCAGTGTCTGATATAAATCTTGAAGAAAGTAAACCAA 1796
Db 1990 GCATATCTGAACATAGTTCAATCAATGCTACTGATAATAAATCTGCTGAAGAAATTAACCAA 2049
QY 1797 GTGAGTTGCAAG 1856
Db 2050 ATCAGGTTGGAG 2109
QY 1857 GGGAGGAGTGTCTCAGCGGAGTCTCTCGGCTTTCAGAGAGAGAGAGAGAGAGAGAGAG 1916
Db 2110 AGGAGGTTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2169
QY 1917 ACTTACCTTAAATGTGCGGCTGAGTGTCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 1968
Db 2170 ACTTACCTTACAGATGAGAGTGGAAATTTTTCATTAATGTTTTCATCAAGTTTTCAT 2229
QY 1969 ---AGTTGACACAGGTTGATAGCATAAAGTCACTGCTCT 2003
Db 2230 CTAAGTTGAAACAGGTTGTGTGATAAAGTCAAGTTTATCT 2267

RESULT 2

US-09-513-999C-3004
; Sequence 3004, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.Y.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 3004
; LENGTH: 428
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS

NO PAGE LEFT BLANK

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 12, 2006, 19:14:34 ; Search time 51 Seconds
(without alignments)
1055.516 Million cell updates/sec

Title: US-10-644-084-2
Perfect score: 3165
Sequence: 1 MGDWMTVDPVLCTENKNLS.....CYSCSSAFSAHGRRDLP 615

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /EMC_Celerra_IDS3/ptodata/2/iaa/5_COMB.pap:*
- 2: /EMC_Celerra_IDS3/ptodata/2/iaa/6_COMB.pap:*
- 3: /EMC_Celerra_IDS3/ptodata/2/iaa/7_COMB.pap:*
- 4: /EMC_Celerra_IDS3/ptodata/2/iaa/H_COMB.pap:*
- 5: /EMC_Celerra_IDS3/ptodata/2/iaa/PCTUS_COMB.pap:*
- 6: /EMC_Celerra_IDS3/ptodata/2/iaa/RE_COMB.pap:*
- 7: /EMC_Celerra_IDS3/ptodata/2/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2152.5	68.0	504	2	US-10-104-047-3467
2	612.5	19.4	137	2	US-09-513-999C-7081
3	223.5	7.1	3259	2	US-09-949-016-6507
4	222.5	7.0	905	2	US-09-248-796A-16333
5	222.5	7.0	2733	2	US-09-949-016-11433
6	220.5	7.0	1780	2	US-09-949-016-6899
7	220.5	7.0	1786	2	US-09-949-016-7880
8	215.5	6.8	2125	2	US-09-913-172-29
9	212.5	6.7	2663	2	US-09-538-092-1252
10	211.5	6.7	1388	1	US-08-685-576-1
11	211.5	6.7	2662	2	US-09-595-684B-31
12	210.5	6.7	1354	2	US-08-685-871-2
13	208.5	6.6	1388	1	US-08-685-576-4
14	208.5	6.6	1388	2	US-09-976-594-296
15	208	6.6	2704	2	US-09-538-092-1260
16	206.5	6.5	1676	2	US-09-949-016-7610
17	206.5	6.5	2186	2	US-09-949-016-10828
18	206.5	6.5	2349	2	US-09-538-092-914
19	204	6.4	1427	2	US-09-538-092-1044
20	203	6.4	3210	2	US-09-538-092-1154
21	203	6.4	3248	1	US-08-353-700-1
22	203	6.4	3248	5	PCT-US95-16216-1
23	200.5	6.3	1958	2	US-10-028-946-4
24	200.5	6.3	2053	2	US-09-964-956-11
25	200.5	6.3	2054	2	US-10-028-946-2
26	200.5	6.3	2066	2	US-09-964-956-9

ALIGNMENTS

RESULT 1

US-10-104-047-3467
; Sequence 3467, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cdna
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104.047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3467
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3467

Query Match	68.0%;	Score	2152.5;	DB 2;	Length	504;	
Best Local Similarity	90.3%;	Pred. No.	9.2e-157;				
Matches	419;	Conservative	20;	Mismatches	5;	Gaps	2;
Qy	1	MGDWMVTDPVLCTENKNLSQYTSKSPSSLYSQOVLCSVPKSNVHGVFGVCTGE	60				
Db	1	MGDWMVTDP-----BSKTSQYTSKSPSSLYSQOVLCSVPKSNVHSPFSFACTED	56				
Qy	61	NIEQISYILODELTFPGPSLYEESKREKRELINIVAVLNCMKNELLVLQRKNLLAQESV	120				
Db	57	NIEQISYILODELTFTFGPFPSLYEESKRGKRELINIVAVLNCMKNELLVLQRKNLLAQENV	116				
Qy	121	ETQNLKLGSDMDHLQSCYAKLEQLETSRREMIGLQERDRLOQCNKRSIHLCLKNEKDEV	180				
Db	117	ETQNLKPGSDMDHLQSCYSKLEQLETSRREMIGLQERDRLOQCNRLHQLLKNEKDEV	176				
Qy	181	OKLQNIITASRATQYNDHVKREREYNYKLERHOLVMNKKDKNIAMDVLNTYVGRADKRG	240				
Db	177	OKLQNIITASRATQYNDHMKREREYNYKLERHOLVMNKKDKNIAMDVLNTYVGRADKRG	236				
Qy	241	SWRTDKTEARNEEDMYKILLNDVEYRQKQILMENAEKVKVLOOMKEMISLLSPQKKKPR	300				
Db	237	SWRTGKTEARNEEDMYKILLNDVEYRQKQILMENAEKVKVLOOMKEMISLLSPQKKKPR	296				
Qy	301	ERAECDGTGTVASIEDDSGELSRDSVMGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ	360				
Db	297	ERVDDSTGTV-ISDVEEDAGELSRDSNMDLSCTVREQLTNSIRKQWRILKSHVEKLDNQ	355				
Qy	361	ASKVHSEGLNEEDVISRQDHEQTEKLELEIRCKEMIKAAQQQLLOQLATTCDDDTTSL	420				

THIS PAGE LEFT BLANK

Db 356 VSKVHLGFGNDEVISRDHEQETEKLELEIOCKKEMIKTQOOLQOQLATAYDDDTSL 415
Qy 421 LRDYILLEEKERLKEEWTLPEKOKKFERRRSFTAAIRLGL 464
Db 416 LRDYILLEEKERLKEEWSLPEKOKKFERRRSFTAAIRLGL 459

RESULT 2

US-09-513-999C-7081
; Sequence 7081, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 7081
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-513-999C-7081

Query Match 19.4%; Score 612.5; DB 2; Length 137;
Best Local Similarity 87.0%; Pred. No. 1.8e-39;
Matches 120; Conservative 10; Mismatches 7; Indels 1; Gaps 1;
Qy 272 MENAELKVLQOMKEMISLLSPKKPRERAEDGTGTVAISDIEDSGELSDVWGLS 331
Db 1 MENAELKVLQOMKEMISLLSPKKPRERVDSTGTV-ISDVEDAGELSDRESMWDL 59
Qy 332 CDTVREQLTNSIRKQWILKSHVEKLNOASKVHSEGLNEEDVTSRODHQETEKLELEI 391
Db 60 CETVREQLTNSIRKQWILKSHVEKLNOASKVHSEGLNEEDVTSRODHQETEKLELEI 119
Qy 392 ERCKEMIKAOQLLQOQL 409
Db 120 QOCKEMIKTQOQLLQOQL 137

RESULT 3

US-09-949-016-6507
; Sequence 6507, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6507
; LENGTH: 3259
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6507

Query Match 7.1%; Score 223.5; DB 2; Length 3259;
Best Local Similarity 19.9%; Pred. No. 1e-07;
Matches 147; Conservative 136; Mismatches 218; Indels 237; Gaps 34;
Qy 15 ENKNLSQVTSSTKMS-----PSSLYSQOVLCSVP-LSKNVGHVGVFCTGENTBQ 64
Db 1235 ENENIGDQLRQIQVRESIDGKLPT--DQESCSSTPGLPEPLFKATEQHHTQVLES 1282
Qy 65 SI-----SYLDOELTTFGPPSLYE---ESKSEAKRELNIIVAVLNMHELLVLQRKNLLA 116
Db 1283 NLCPDWPSHSEDASALQGTSTVAQIKAQIKETAEAKVELELKVSTTSELTKKSEVFPOL 1342
Qy 117 QESVETONLKLGS--DMDHLOSCYAK-LKEOELSRREMIGLOERDROLOCKNRSILHQLL 173
Db 1343 QEQINKQGLEIESLKTVSHEAEVHAESLQCKLESSQLQIAGL-EHLRELQPK----- 1393
Qy 174 KNEKDEVOKLQNIIASRATQYNHDVKKREYNKLERLHQLVMNKKDKNIAMVNLVYVG 233
Db 1394 ---LDELQKL-----ISKKEEDVSLGQL-----SEKEAALTKI----- 1425
Qy 234 RADGRGSRWRTDKTARNEDENYKIL-----LNDYEVROKQILMENAELK----- 278
Db 1426 -----QTEIIQEDILIKALHTOLEMOAKEHDERIKOLOVELCEMKQKPEIGR 1473
Qy 279 --KVLOOMKEMISLLSPOK-----KKPRERAEDGTGTV-----AISDIED--DSGELS 323
Db 1474 ESRKQOIQRKLOALISRKALKENKSLQBELSLARGTIERLTSLADVESQVSAQNK 1533
Qy 324 RDSVMG-----LSCDTVR-----EQL 339
Db 1534 KUTVLGRLLQOERDKLITENDRSILENQSLSSCSLKLALGLETKELVKIESL 1593
Qy 340 TNS-----IRKQWILKSHVEKLNOASKVHS--EGLNEE-----DVISR 377
Db 1594 KSKIAESTEWQCKELQKEYEILLQSYENVSNEAERIOHVVEAVROEKQELYGKLRT 1653
Qy 378 QDHEQETEK---LELEIERCKEMIK---AQOQLLQOQ-----LATTTC 413
Db 1654 EANKKETEKLOQEAQOEMEMKEMKPKFPAKSKQOQKILELEENDRLRAEVHPAGDTAKEC 1713
Qy 414 DDDTTSLLRDCVLLBEKERLKEEWTLPEKOKKFERRRSFTAAIRLGLERKAFEBERA 473
Db 1714 ME---TLLSSNASMKEELERVKMEYETLSKKFQSLMSEKDSLSEEVQDL-----KHQIEDNV 1767
Qy 474 SVWQOFLNMTNPDHONSENVK---LFSAFSGSSDDNLIVHSRPRQKLSHVSANGVPAC 530
Db 1768 S--KQANLEATE-KHDNQTNTVEGTQSTPGTETEQDSLMSSTR-----PTC 1811
Qy 531 TSKLTKSLPASSTS-DFRQTHSCVSEHSSIS-----VLNITPEESKPSSEVAR 577
Db 1812 SESVPSAKSANPAVSKDF-----SSHDEINNYLQOIQIDOLKERIAGLBEKQKNKEPSQ 1864
Qy 578 ESTDQKWSQSRPSSREG 595
Db 1865 TLENERNTLLSQSLSTKDG 1882

RESULT 4

US-09-248-796A-16333
; Sequence 16333, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208

THIS PAGE LEFT BLANK

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 13, 2006, 23:29:14 ; Search time 304 Seconds
(without alignments)
5677.954 Million cell updates/sec

Title: US-10-644-084-2
Perfect score: 3165
Sequence: 1 MGDWMTVPVLTENKRLS.....CYSGCSAFRSHGRDRLP 615

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/abes/ABESWEB_spool/US10644084/runat_12062006_150424_1461/app_query.fasta_1
-DB=Issued Patents NA -QFW=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=Bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-HOST=abs03p -USER=US10644084 @CEN 1.1 307 @runat_12062006_150424_1461
-NCPUs=6 -ICPU=3 -NO MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRVTIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

1: /EMC_Celestra_SIDS3/ptodata/2/ina/1 COMB.seq:
2: /EMC_Celestra_SIDS3/ptodata/2/ina/5 COMB.seq:
3: /EMC_Celestra_SIDS3/ptodata/2/ina/6A COMB.seq:
4: /EMC_Celestra_SIDS3/ptodata/2/ina/6B COMB.seq:
5: /EMC_Celestra_SIDS3/ptodata/2/ina/7 COMB.seq:
6: /EMC_Celestra_SIDS3/ptodata/2/ina/H COMB.seq:
7: /EMC_Celestra_SIDS3/ptodata/2/ina/FCUS COMB.seq:
8: /EMC_Celestra_SIDS3/ptodata/2/ina/PP COMB.seq:
9: /EMC_Celestra_SIDS3/ptodata/2/ina/RE COMB.seq:
10: /EMC_Celestra_SIDS3/ptodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2720	85.9	2272	3	US-10-104-047-1497 Sequence 1497, Ap
2	635.5	20.1	428	3	US-09-513-999C-3004 Sequence 3004, Ap
3	223.5	7.1	10300	3	US-09-949-016-636 Sequence 636, App
4	222.5	7.0	2718	3	US-09-248-796A-2230 Sequence 2230, Ap
5	222.5	7.0	8590	3	US-09-949-016-5562 Sequence 5562, Ap
6	222.5	7.0	38575	3	US-09-949-016-17304 Sequence 17304, A
7	222.5	7.0	119153	3	US-09-949-016-12378 Sequence 12378, A
8	220.5	7.0	6276	3	US-09-949-016-2009 Sequence 2009, Ap

9	220.5	7.0	6284	3	US-09-949-016-1028 Sequence 1028, Ap
10	217	6.9	6921	3	US-09-643-597-117 Sequence 117, App
11	217	6.9	6921	3	US-09-480-884A-117 Sequence 117, App
12	217	6.9	6921	3	US-09-542-615A-117 Sequence 117, App
13	217	6.9	6921	3	US-09-606-421B-117 Sequence 117, App
14	217	6.9	6921	3	US-09-221-107-117 Sequence 117, App
15	217	6.9	6921	3	US-09-466-396A-117 Sequence 117, App
16	217	6.9	6921	3	US-09-476-496A-117 Sequence 117, App
17	217	6.9	6921	3	US-09-630-940B-117 Sequence 117, App
18	217	6.9	6921	3	US-09-285-479-117 Sequence 117, App
19	217	6.9	6921	3	US-10-007-700-117 Sequence 117, App
20	217	6.9	7045	3	US-09-919-172-28 Sequence 28, Appl
21	215	6.8	8930	3	US-09-814-915A-91 Sequence 91, Appl
22	212.5	6.7	8257	3	US-09-595-684B-30 Sequence 30, Appl
23	212.5	6.7	8503	2	US-09-620-312D-130 Sequence 130, App
24	211.5	6.7	5053	2	US-08-685-576-2 Sequence 2, Appl
25	210.5	6.7	4065	3	US-09-016-434-1105 Sequence 1105, Ap
26	210.5	6.7	4739	3	US-08-685-871-1 Sequence 1, Appl
27	209.5	6.6	5857	3	US-09-220-132-79 Sequence 79, Appl
28	208.5	6.6	4363	2	US-08-685-576-5 Sequence 5, Appl
29	208.5	6.6	4848	3	US-09-976-594-295 Sequence 295, App
30	208	6.6	2728	3	US-09-799-451-62 Sequence 62, Appl
31	206.5	6.5	5813	3	US-09-949-016-1739 Sequence 1739, Ap
32	206.5	6.5	7500	3	US-09-949-016-4957 Sequence 4957, Ap
33	203.5	6.4	6481	3	US-09-620-312D-1076 Sequence 1076, Ap
34	203	6.4	4747	3	US-09-854-133-729 Sequence 729, App
35	203	6.4	7218	3	US-09-949-016-1775 Sequence 1775, Ap
36	203	6.4	7218	3	US-09-949-016-1776 Sequence 1776, Ap
37	203	6.4	10136	2	US-08-353-700-2 Sequence 2, Appl
38	203	6.4	10136	7	PCT-US95-16216-2 Sequence 2, Appl
39	203	6.4	82125	3	US-09-949-016-13517 Sequence 13517, A
40	203	6.4	82125	3	US-09-949-016-13518 Sequence 13518, A
41	202	6.4	5877	3	US-10-028-946-3 Sequence 3, Appl
42	202	6.4	6165	3	US-10-028-946-1 Sequence 1, Appl
43	202	6.4	6189	3	US-09-964-956-10 Sequence 10, Appl
44	202	6.4	6201	3	US-09-964-956-8 Sequence 8, Appl
45	201	6.4	6159	3	US-10-017-216-3 Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-10-104-047-1497
; Sequence 1497, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1497
; LENGTH: 2272
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1497

Alignment Scores:
Pred. No.: 2.48e-284 Length: 2272
Score: 2720.00 Matches: 536
Percent Similarity: 86.8% Conservative: 34
Best Local Similarity: 81.6% Mismatches: 40
Query Match: 85.9% Indels: 47
DB: 3 Gaps: 3

US-10-644-084-2 (1-615) x US-10-104-047-1497 (1-2272)

Qy 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
|||||
|||||

THIS PAGE LEFT BLANK

87

224 ATGGGAGATTGGATGACTGTTTACAGATCCA-----GAAAGCAAAATATCTCT 271
Db
21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuLys 40
Qy
272 CAATATACCTCAGAAACAAAGATGCTCCATCAAGTTTATCTACTCAGCAAGTGTATGT 331
Db
41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Qy
332 TCTTCATACCTTTATCGAANAATGTGCACAGTTTTTTTCAGTGCCTTCGCACAGAAGAT 391
Db
61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGlnLeuThrThrPheGlyPheProSer 80
Qy
392 AATATTGAACAGAGTATCTCATCTCTGATCAGAAATGACTACTTTTGGTTTTTCCCTCA 451
Db
81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Qy
452 TTATATGAAGAAATCCAAAGGTAAAGACAAGAGAGAGTTAATATATAGTAGTGTACTA 511
Db
101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerVal 120
Qy
512 AATTGTATGATGAGTCTGTTGCTTCAGCGGAAGAACCTTCTAGCTCAGCAAAATGTG 571
Db
121 GluThrGlnAsnLeuLysGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
Qy
572 GAGACACAGAATTTGAGCGCGGAAGTATATGACCACTTACAGAGCTGCTACTCAAAA 631
Db
141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGlnLysArg 160
Qy
632 CTTAAGGAACAACCTGGAAACCTCCAGGAGGGAATGATGGCTTCAGGAAGAGACAGA 691
Db
161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
Qy
692 CAGTTACAAATGTAAGACAAGCAAGTATGTCATCAGCTACTTAAAGAAATGAGAAAGATGAGGTG 751
Db
181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Qy
752 CAAAATATACAAATATATCTGCAAGTCGAGCTACTCAGTATATATGATATGAAGAGA 811
Db
201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Qy
812 AAGAGCGGTGAATATAAATGAACTGAAGACGCTCATCACTTGTATGACAGAA 871
Db
221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Qy
872 GATAGAAAATAGCTATGACATTTTGAATATTGTCGAGAGCTGATGGAAGAAAGAGGC 931
Db
241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrIleLeuLeu 260
Qy
932 TCCTGGAGGACTGGTAAACCTGAAGCCAGGAATGAAGATGAATGTATAAATTTCTCTTG 991
Db
261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
Qy
992 AATGATTATGAATATCGTCGAACAAATCTTAATGGAATATGCAAGAACTTAAGAGGTT 1051
Db
281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
Qy
1052 CTTCAACAAATGAAAGAAATGATTTCTCTTCTTCTCCCAAAAGAAAGAACTTACA 1111
Db
301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
Qy
1112 GAAAGAGTAGATGATAGTACAGGAACCTGTT---ATTTCGATGTTGAAGAAAGTCCCGGG 1169
Db
321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
Qy
1169 GAACATAGCAGAGAGATGATGGGACCTTTCTGTGAACCTGTGAGAGAGCAGCTTACA 1228
Db
341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
Qy
1229 AACAGCATCAGAAAACAGTGGAGAAATTTTGAAGATCATGTAGAAAAGCTTCAATACCAA 1288
Db
361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
Qy
1289 GTTTCAAAGGTACACCTGGAAAGGTTTTTAATGATGAGATGTAATCTCACGCAAGACCAT 1348
Db

381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400
Qy
1349 GAACAGAAACTGAAAAACTCGAGTTAGAAATTCAGCAGCTGTAAAGAAATGATTAAACT 1408
Db
401 GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
Qy
1409 CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCTATGATGATGATGATGATGATGAT 1468
Db
421 LeuArgAspCysTyrLeuLeuGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440
Qy
1469 TTACGAGAGCTGTTATTTGTTGGAAGAAAGAACGCTCTCAAGAGAAATGTCCTCTTTT 1528
Db
441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
Qy
1529 AAAGAGCAGAAAAGAAATTTTGAGAGGAGAGAGCAAGCTTTTACAGAAGCCGCTATTTCGC 1588
Db
461 LeuGlyLeuGlu----- 464
Qy
1589 CTGGGATTTGAGATTTGCTTCTCTGCCAGTGTTTTACAGAAGAAATCTCTTTTGGGT 1648
Db
464 ----- 464
Qy
1649 ATCCATCATTCATCCACGGTCAGTAGTGTGTGCAGAGACAGTCGCAACCAAGACTAT 1708
Db
465 -----ArgLysAlaPheGluGluArgAlaSerTrpValLysGlnG 479
Qy
1709 ATTTGCTGTTTAAAGAAAGCAATTTGAAGAGAAAGACCCAGTTGGTTAAAGCAGCA 1768
Db
479 nPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAl 499
Qy
1769 GTTTCTAAATATGACTACTTTGACCAACCAAGACTCAGAAAATGTGAAACTTTTTCAGTGC 1828
Db
499 aPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLy 519
Qy
1829 CTTCTCAGGAAGTCTGATTGGGACAATCTTATAGTCACCTCGAGGCGCGCAAAAGAA 1888
Db
519 sLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuPr 539
Qy
1889 GCCTCAGAGTGTCTAATGGGTCTCCAGTTTGCATGCTCTAAACTTAACTTAACTCTTCC 1948
Db
539 oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe 559
Qy
1949 TGCTTCACTTCCACTTCAGACTTTTCCAGACAGCAGCTTCTGTCATATCTGACATAGTTC 2008
Db
559 rIleSerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSe 579
Qy
2009 AATCAATGTAAGTAAATATACTGCTGAAGAAATTAACCAAAATCAGGTTCGAGGAGAACG 2068
Db
579 rThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerG1 599
Qy
2069 TACAATCAAAAATGGAGTGTGGCTCAGACCTGGATCACAGGAAGGTTGCTATAGTGG 2128
Db
599 yCysSerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615
Qy
2129 ATGCTCTTGGAGCTACACAAATCTCATGTAGAAAAAGATGACTTACCT 2177
Db

RESULT 2

US-09-513-999C-3004
; Sequence 3004, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Ductert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59 US2. REG
; CURRENT APPLICATION NUMBER: US/09/513.999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122.487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681

THIS PAGE LEFT BLANK

Qy	1981	GTGTAGCATAAAGTCAGTCGTTCTAACTTAAAGATGCTCAGAGTTGTTTGGATTTCGC	2040
Db	1981	GTGTAGCATAAAGTCAGTCGTTCTAACTTAAAGATGCTCAGAGTTGTTTGGATTTCGC	2040
Qy	2041	TGTCCTTCCCCCAAGAGCTGAAATGCTAAGCTACTTAAAGAGATGCAAAAGCTTTGGTTGT	2100
Db	2041	TGTCCTTCCCCCAAGAGCTGAAATGCTAAGCTACTTAAAGAGATGCAAAAGCTTTGGTTGT	2100
Qy	2101	GTGTTAGTAAACAGAGCCCCCTGGCTCTGTGTGACTCAGAGAAATGCAATGCGTTTGGATGAA	2160
Db	2101	GTGTTAGTAAACAGAGCCCCCTGGCTCTGTGTGACTCAGAGAAATGCAATGCGTTTGGATGAA	2160
Qy	2161	ACAGAAGCGCTGGAATGATTTGCCCTCGCCAGGTACCGAAGAGACATTTTAGGGACTGGT	2220
Db	2161	ACAGAAGCGCTGGAATGATTTGCCCTCGCCAGGTACCGAAGAGACATTTTAGGGACTGGT	2220
Qy	2221	TCCTGTAAACATTAATAATTTCCGTCCCAAGTGTGGTTCGCAATGGAAAGTGTAGCCCTTACT	2280
Db	2221	TCCTGTAAACATTAATAATTTCCGTCCCAAGTGTGGTTCGCAATGGAAAGTGTAGCCCTTACT	2280
Qy	2281	TGAATGTATACTGTAGATTTTTTAAACAAAGCAGGTTCTATATTTATTATGTTTGTAGTGTAT	2340
Db	2281	TGAATGTATACTGTAGATTTTTTAAACAAAGCAGGTTCTATATTTATTATGTTTGTAGTGTAT	2340
Qy	2341	TTTCGGGATTAACCTCTTTTCATATGTTTGTGTCCTGTACATAAAATATACATGATGTTAA	2400
Db	2341	TTTCGGGATTAACCTCTTTTCATATGTTTGTGTCCTGTACATAAAATATACATGATGTTAA	2400
Qy	2401	GAGGCTTTTAAAGGTTTAAAAACCTTCACACCATGCTTGAGTAGTATAGCATTTTCATGCCAATTAA	2460
Db	2401	GAGGCTTTTAAAGGTTTAAAAACCTTCACACCATGCTTGAGTAGTATAGCATTTTCATGCCAATTAA	2460
Qy	2461	AATGTTTTTCAGTGCGATGGTGTTTACAGAGGTTAGGACCACTGCCACATGACATGTTAAGA	2520
Db	2461	AATGTTTTTCAGTGCGATGGTGTTTACAGAGGTTAGGACCACTGCCACATGACATGTTAAGA	2520
Qy	2521	CTTTATTTTTTAAAGCACTCGGGCAATAAAATTTCAAGGCCCTTCATTAAGCTGAGTTTCAG	2580
Db	2521	CTTTATTTTTTAAAGCACTCGGGCAATAAAATTTCAAGGCCCTTCATTAAGCTGAGTTTCAG	2580
Qy	2581	ATAACTAGAACTACTAAACGTTTACATTTTTCAGATTTTTTAAAGCAATGTTATTTTATTTTAT	2640
Db	2581	ATAACTAGAACTACTAAACGTTTACATTTTTCAGATTTTTTAAAGCAATGTTATTTTATTTTAT	2640
Qy	2641	ATATGTGAATGTTATAATTTCTAAGAGGAATAATGATTATGGAGTAATGGG	2692
Db	2641	ATATGTGAATGTTATAATTTCTAAGAGGAATAATGATTATGGAGTAATGGG	2692

RESULT 2	
ADO26492	
ID	ADO26492 standard; DNA; 3195 BP.
XX	
AC	ADO26492;
XX	
DT	29-JUL-2004 (first entry)
XX	
DE	Rat afadin dilution domain binding protein (ADIP) DNA SeqID 3.
XX	
KW	rat; gene; ds; afadin dilution domain binding protein; ADIP; afadin;
KW	actinin; binding inhibitor; cardiant; heat disease;
KW	myocardial infarction; myocarditis.
XX	
OS	Rattus norvegicus.
XX	
FH	Location/Qualifiers
FT	79.1920
CDS	
FT	/*tag= a
FT	/product= "ADIP protein"
XX	
XX	JP2004135658-A.
XX	

13-MAY-2004.

14-AUG-2003; 2003JP-00293554.

27-SEP-2002; 2002JP-00284263.

(EISA) EISAI CO LTD.

WPI; 2004-404616/38.

P-PSDB; ADO26493.

New polynucleotide encoding an afadin dilution domain binding protein having avidity with afadin or actinin, useful for diagnosing heart diseases e.g. myocardial infarction.

Claim 1; SEQ ID NO 3; 37pp; Japanese.

This invention relates to a novel isolated nucleic acid encoding an afadin dilution domain binding protein (ADIP) that exhibits an avidity with afadin/actinin. Specifically, it refers to screening assays to identify compounds that modulate ADIP avidity and provides suitable agonists, antagonists and antibodies thereof. The present invention provides methods to identify afadin and actinin binding inhibitors therapeutically as cardiants to diagnose and/ or treat heart disease such as myocardial infarction or myocarditis. This polynucleotide sequence is the rat ADIP DNA sequence of the invention.

Sequence 3195 BP; 929 A; 648 C; 794 G; 823 T; 0 U; 1 Other;

Query Match 81.3%; Score 2189, 4; DB 12; Length 3195;
Best Local Similarity 90.1%; Pred. No. 0;
Matches 242; Conservative 0; Mismatches 247; Indels 20; Gaps 7;

Qy 2 GTAGGAGAGTGACAGGAGCTGTGTGAAGCGTCGCAGCAGCTCAGCGCGCTCTCAGGTATC 61
Db 1 GTAGGAGAGTGACAGGAGCTGTGTGCATGCCCCAGCACCTGAATGCTTCTCAGGGACC 60

Qy 62 CTGGCTCTGGAACCTTGCTATGGAGATTGATGACTGTGACAGATTCAGTTCGTGTGACA 121
Db 61 CTGGCTCTGGGACTGGCTATGGAGATTGGATGACTGTGTTACAGATTCAGTTCGTGTGACA 120

Qy 122 GAAAAACAAAATCTCTCAATATACCTCAGAAACAAAAGATGTCTCGTCCAGTTTGTGAC 181
Db 121 GAAAAACAAAATCTCTCAATATACCTCAGAAACAAAAGATGTCTCGTCAAGTTTATAC 180

Qy 182 TCCAGCAAGTTCGTGCTCTTCAGTACCTTTATCCAAAAACGTGATGTTGTTTCGGT 241
Db 181 TCGCAGCAAGTACTGTGCTCTGCAACACCTTTATCCAAAGAAATGTGCATGTGTTTTCAGT 240

Qy 242 GTCTTCTGCACAGAGAGAACATTGAAACAAGATATTTCCTATCTTGATCAGGAGCTGACC 301
Db 241 GCCTTCTGCACAGAGAGAACATCGAACAGAGATTTTCGTATCTTGATCAGGAACTGACT 300

Qy 302 ACCTTCGGGTTTCCTTCTTGATGAAGAAATCCAAAAGTAAAGAGGCAAAAGAGAGAATTA 361
Db 301 ACCTTCGGTTTCCTTCTTGATGAAGAAATCCAAAAGTAAAGAGGCAAAAGAGAGTTA 360

Qy 362 AATATAGTCGCTGTTCTGAACCTGATGAACGAGCTGCTCGTGTCTTCAGCGGAAGAACCTG 421
Db 361 AGTATAGTTGCTCTTCTGAACCTGATGAATGAGCTGCTGTGTGCTTCAGCGGAAGAACCTC 420

Qy 422 CTGGCCCCAGGAGAGCTGGAGACACGAACTTGAAGCTGGGCACTGACATGGACCCACCTG 481
Db 421 CTGGCCCCAGGAAGCTGGAGACACGAAATCTGAAGCTGGGCACTGACATGGACCCACCTG 480

Qy 482 CAGAGCTGCTACGCCAAAACCTTAAGGAGCAGTTTGGAAAACGTTCAGCGCGGAGATGATCGGG 541
Db 481 CAGAGCTGCTACGCCAAAACCTTAAGGAAACAGTTTGGAGGCCCTCCAGGGCGAGAGATGATCAGC 540

Qy 542 CTTCAAGAGAGACAGGAGCTGCAAGTGCAGAAACAGGAGTTTGGATCAGCTCCTGGAAG 601
Db 541 CTTCAAGAGAGACAGAGCTGCAAGTGCAGAAACAGGAAATTTGGATCAGCTCCTGGAAG 600

Qy	602	AATGAGAAAGATGAGGTACAAAATATCAAAATATCATAGCCAGCGCGCTACTCAGTAT	661
Db	601		
Qy	662	AATCATGATGTGAAGAGGAGCGTGAATATATTAAGCTTAAAGGAGCGCTGCATCAG	721
Db	661		
Qy	722	CTCGTTATGAACAAGAGGATAAAAATATAGCCATGGATGTTTTAAATATATGCGGTGGA	781
Db	721		
Qy	782	GCTGATGCAACAGAGGCTCATGAGGACTGACAAAACAGAAAGCAGGAATGAAGATGAG	841
Db	781		
Qy	842	ATGTACAAAATCTGTGTAATGATATATGAGTACCGCCAGAAAGCAGATCTCTGCTCCT	901
Db	841		
Qy	902	CGGAGCTGAAGAGGTCCTCCAGCAGATGAAGAGGAGATGATCTCTCTCTCTCCT	961
Db	901		
Qy	962	CAGAAGAAAGCCAGGGAAGAGCAGAGGACGCGCAGGCACCTGTTGCTATCTCCGAT	1021
Db	961		
Qy	1022	ATAGAAGATGACTCTGGGGAACCTGAGCAGAGACAGCGGTGGGCGCTTCTCTGACACT	1081
Db	1021		
Qy	1082	GTGAGAGCAGCTGACAAAACGATCAGGAAAACGTTGAGAGATTTTGAAAGTCATGTA	1141
Db	1081		
Qy	1142	GAAAACTCGATTAACCAAGCTTGAAGGTACACTCAGAGGCGCTTAATGAGGAGGAGCTC	1201
Db	1141		
Qy	1202	ATCTCAGCAAGACCATGAGCAAGAGACTGAGAAACTGGAGCTGGAGATTTGACGGGTG	1261
Db	1201		
Qy	1262	AAAGAGATGATCAAGGCTCAGCAGAGCTCTTACAGCAGAGCTGGCCACACCTGTGAT	1321
Db	1261		
Qy	1322	GATGACACCACTCACTGTCGAGACTGTTACTTGTGGAGAAAAGGAAAGGAAAGCGCTTAA	1381
Db	1321		
Qy	1382	GAAGAGTGGACCCCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAAAGCGGAAGCTTT	1441
Db	1381		
Qy	1442	ACAGAACTGCCATTCGATTTGGGTTGGAGAGAAAGCGCTTTGAAGAGCGGAGCCAGC	1501
Db	1441		
Qy	1502	TGGGTAAAGCAGCAGTGTTTTAAACATGACGAACCTTTGACCACAGAACTCAGAAAATGTG	1561
Db	1501		
Qy	1562	AAACTTTTTCAGTGCCTCTCAGGAAGTTCTGATCCAGACAATCTTATAGTCCACTCAGG	1621
Db	1561		
Qy	1622	CCAGCGAAAAGNAGCTACACAGTGTGGCTAATGGGTGCCAGCTTGCACATCAAACTG	1681
Db	1621		
Qy	1682	ACTAAATCTCTCTCGCTCACCCTTCTACTTCAGACTTTTCGCCAGACACATTTCTGTGTG	1741

Db	1681		
Qy	1742	TCTGAACACAGTTCATCAGTGTGCTGAATATTAACCTCCTGAAGAAAGTAAACCAAGTGAG	1801
Db	1735		
Qy	1802	GTTTGAAGAGAAAGCAGGATCAGAGTGGAGCGTGAGTCGAGGCCAGCTCCCGGAG	1861
Db	1795		
Qy	1862	GGGTGCTTACAGCGGATGCTCCTCGGCCTTCAGGAGCGCTCAACCGGAGCCGAGATGACTTA	1921
Db	1855		
Qy	1922	CCTTAAATGTGCGGCTGCGAGTGTCTCCAGATGTGCGCTAGAGAGTTGACACAGG	1981
Db	1915		
Qy	1982	TGTAGCATAAAGTCAGTCGCTAACTTAAAGATGCTCAGAGTTGTTCTGTTGGACTTCGCT	2041
Db	1973		
Qy	2042	GTCTTCCCCAAAAGAGCTGAAATGCTAAGCTACTTAAAGGATGCAAAAGCTTTTGGTGTG	2101
Db	2032		
Qy	2102	TGTTAGTAAACAGAAAGCCCTGGCTCTGTGACTGACAGGAATGCAATGCGGTTTGGATGAAA	2161
Db	2085		
Qy	2162	CAGAAGCGCTGGAATGATTCGCTC-GCCAGGTACCGAGAAAGAGACACTTTTAGGAGCTGGT	2220
Db	2145		
Qy	2221	TCCTGTAACCATTAATTAATTCGTCCCAGTGTGGTTCGCATTTGGAAGTGTAGCTTTACT	2280
Db	2205		
Qy	2281	TGAATGTATACTAGATTTTAAACAAAGCAGGTTCTATATTTATTTATTTAGTGTGAT	2340
Db	2265		
Qy	2341	TTTGGGATTAACCTCTTTTCATATGTTTGTCTGTACATAAATATACATGACTATGTAA	2400
Db	2324		
Qy	2401	GAGCTTTAAGGTTTAAAACTTCAACCATGCTTGAAGTGTAGCATTTTCATGCCAAATTA	2460
Db	2384		
Qy	2461	AATGTTTTTCAGTGCATGGTGTTTACAGA--GGTTAGGACCACTGCCACATGACAGTTAA	2518
Db	2444		
Qy	2519	GACTTTATTTTAAAGCCATCTGGGCAATAAAAAATTTCAAAGCCCTTTCATAGCTGAGTTC	2578
Db	2504		
Qy	2579	AGATAACTAGAACTACTAAACGTTTACATTTTGAAGATTTTAAAGCATTTTATTTATTTT	2638
Db	2564		
Qy	2639	ATATATGTAATGTTATTAATTTCTTAAGAGGAATATTCATTTATGAGTAATCGGG	2692
Db	2624		

RESULT 3
ACD19444
ID ACD19444 standard; cdna; 5476 BP.
XX
AC ACD19444;

XX	25-AUG-2003	(first entry)	CC	haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).
XX	CDNA encoding novel human protein #124.	CC	These are also useful in developing powerful assay system for functional	
DE	Human; NOV; gene therapy; endocrine related disease; diabetes;	CC	analysis of various human disorders, as well as in diagnostic	
XX	metabolism-related disease; obesity; central nervous system disorder;	CC	applications, and for monitoring the effects of drugs during clinical	
KW	Alzheimer's disease; Parkinson's disease; epilepsy; multiple sclerosis;	XX	trials. This sequence encodes a novel human NOV protein	
KW	schizophrenia; depression; autoimmune disorder; inflammatory disorder;	SQ	Sequence 5476 BP; 1806 A; 887 C; 1063 G; 1720 T; 0 U; 0 Other;	
KW	psoriasis; allergy; lupus erythematosus; asthma; cancer;	Query Match	60.3%; Score 1624.4; DB 10; Length 5476;	
KW	inflammatory bowel disease; rheumatoid arthritis; osteoarthritis;	Best Local Similarity	80.5%; Pred. No. 0;	
KW	colon cancer; lung cancer; liver cancer; breast cancer; ovarian cancer;	Matches 2128; Conservative	0; Mismatches 456; Indels 60; Gaps 17;	
KW	prostate cancer; brain cancer; melanoma; liver disease; liver cirrhosis;	Qy	102 CAGATCCAGTTCGTGTACAGAAAAACAAAATCTCTCAATATACCTCAGAAAAACAAGA 161	
KW	lung disease; emphysema; obstructive pulmonary disease; haemophilia;	Db	107 CCGCTCCCGGGGCGGTGACGAAAGCAAAACTATCTCTCAATATACCTCAGAAAAACAAGA 166	
XX	stroke; infection; gene; ss.	Qy	162 TGTCTCCGTCAGTTTGTACTCCACCAAGTTCTGTGCTCTTCAGTACCTTTATCCAAA 221	
OS	Homo sapiens.	Db	167 TGTCTCCATCAAGTTTATACTACAGCAAGTGTCTATGTTCTTCATACCTTTATCGAAA 226	
XX	WO2003023002-A2.	Qy	222 ACGTGCATGGTGTTCGGTGTCTTCTGCACAGGAGAGAACATTGAACAAAGTATTTTCCT 281	
XX	20-MAR-2003.	Db	227 ATGTGCACAGTTTTTTTCAGTGCCTTCTGCACAGAGATAATATTGAACAGAGTATCTCAT 286	
PF	09-SEP-2002; 2002WO-US028539.	Qy	282 ATCTTGATCAGGAGCTGACCACTTCGGGTTTCCTTCTGTGTATGAAGAAATCCAAAAGTA 341	
XX	07-SEP-2001; 2001US-0318120P.	Db	287 ATCTTGATCAGGAAATGACTACTTTTTTGGTTTTTCTTCATTATATGAAGAAATCCAAAAGTA 346	
PR	07-SEP-2001; 2001US-0318130P.	Qy	342 AAGAGGCAAGAGAGAGAAATTAATATAGTCGCTGTTCGAACTGTATGAACGAGCTGCTCG 401	
PR	10-SEP-2001; 2001US-0318430P.	Db	347 AAGAGACAAAGAGAGAGATTAAATATAGTAGTGTACTAAATTTGTATGAATGAGTCTGTTG 406	
PR	17-SEP-2001; 2001US-0322636P.	Qy	402 TGCTTCAGCGGAGAAACTCTGCTGGCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGG 461	
PR	17-SEP-2001; 2001US-0322781P.	Db	407 TGCTTCAGCGGAGAAAGAACTTCTAGCTCAGGAAAAATGTGGAGACACAGAAATTTGAAGCTGG 466	
PR	17-SEP-2001; 2001US-0322816P.	Qy	462 GCAGTGACATGGACCACTGCAGAGCTGCTACGCCAACTTAAAGAGGAGCTTGGAAAGCT 521	
PR	17-SEP-2001; 2001US-0322817P.	Db	467 GAAGTGATATGGACCATCTACAGAGCTGCTACTCAAAAACCTTAAAGGAACTTAAAGAAACCT 526	
PR	19-SEP-2001; 2001US-0323519P.	Qy	522 CCAGCGGGAGATGATCGGGCTTCAAGAGAGAGACAGGCGCTGCGAGTGCAGAAACACAGA 581	
PR	20-SEP-2001; 2001US-0323631P.	Db	527 CCAGGAGGGAAATGATTTGGGCTTCAGGAAAGAGACACAGCTTACATGTGAAGAACAGGA 586	
PR	20-SEP-2001; 2001US-0323636P.	Qy	582 GTTTGATCAGCTCCTGAAAGATGAGAAAGATGAGGTACAAAAATTACAAAAATATCATAG 641	
PR	25-SEP-2001; 2001US-0324969P.	Db	587 ATTTGATCAGCTACTTAAAGAAATGAGAAAGATGAGGTGCAAAAATTTACAAAAATATCATTTG 646	
PR	26-SEP-2001; 2001US-0324990P.	Qy	642 CCAGCGGGCTACTCAGTATATCATCATGATGTAAGAGAGAGGAGCGGTGAATATATAAGC 701	
PR	17-APR-2002; 2002US-0373212P.	Db	647 CAAGTCGAGCTACTCAGTATATCATCATGATATGAAGAGAAAGAGCGGTGAATATAAATAAC 706	
XX	06-SEP-2002; 2002US-00236177.	Qy	702 TAAAGAGAGCGCTGCATCAGCTCGTTATGAAACAAGAGAGTAAAAACATAGCCATGGATG 761	
XX	(CURA-) CURAGEN CORP.	Db	707 TGAAGGAACGCTCTACATCAACTTGTATGAACAAGAAAGATGAGAAATAGCTATGAGCA 766	
XX	Spytek KA, Patturajan M, Gorman L, Li L, Anderson DW, Zhong M;	Qy	762 TTTTAAATATATGTTGGTTCGAGCTGATGGCAAAACGAGGCTCATGGAGGACTGCACAAAAACAG 821	
XX	Gerlach VL, Vernet CAM, Ellerman K, Berghs C, Rothenberg ME, Guo X;	Db	767 TTTTGAATATATGTCGGGAGAGCTGATGGAAAAAGAGGCTCTCTGGAGGACTGGTAAAAACTG 826	
PI	Shimkets RA, Leach MD, Catterton E, Kekuda R, Ji W, Miller CE;	Qy	822 AAGCCAGGAATGAAGATGAGATGTACAAAATTCCTGTTGAATGATTATGAGTACCGCCAGA 881	
PI	Rieger DK, Taupier RJ, Shenoy SG, Liu X, Padigar M, Alsobrook JP;	Db	827 AAGCCAGGAATGAAGATGAAATGTATAAAATTTCTTGTGAATGATTATGAATATCGTCAGA 886	
XX	Lepley DM, Edinger SR, Burgess CE;	Qy	882 AGCAGATCCTGATGGAGAACCGGAGCTGAAGAGGTCCTCCAGCAGATGAAGAGAGGAGA 941	
WPI;	2003-313242/30.	Db	887 AACAAATCCTAATGGAATAATGCAGAACTTAAGAAAGTTCTTCAACAAATGAAAAGGAAA 946	
P-PSDB;	ABO14751.	Qy	942 TGATCTCTCTCTCTCTCTCAGAAAGAGACCCCGGAAAGAGACAGAGACGGCACAG 1001	
DR	New cytoplasmic, nuclear membrane bound or secreted polypeptides (NOVX)	Db	947 TGAATTTCTCTCTCTCTCTCTCCCAAAAGAGAAACCTTAGAGAAAGAGTAGATGATAGTACAG 1006	
DR	and polynucleotides, useful in gene therapy, e.g. for treating or	Qy	1002 GCATGTTGCTATCTCCGATATAGAAGATGACTCTCTGGGGAACTGAGCAGACAGACCGTGT 1061	
XX	preventing obesity, multiple sclerosis, allergy, cancers, hemophilia,			
XX	stroke or infections.			
XX	Claim 20; Page 351-353; 586pp; English.			
XX	The invention describes a new isolated polypeptide (NOVX). The NOVX			
CC	polypeptide, nucleic acid and antibody are useful as therapeutics,			
CC	particularly in the manufacture of a medicament for treating a syndrome			
CC	associated with a human disease, which includes a pathology associated			
CC	with NOVX polypeptide. The DNA encoding the protein is useful in gene			
CC	therapy for treating the disease or condition. In particular, the NOVX			
CC	polypeptide or polynucleotide is useful for treating endocrine/			
CC	metabolism-related diseases (e.g. obesity or diabetes), central nervous			
CC	system disorders (e.g. Alzheimer's disease, Parkinson's disease,			
CC	epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune			
CC	and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,			
CC	asthma, inflammatory bowel disease, rheumatoid arthritis or			
CC	osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,			
CC	prostate or brain cancers, or melanoma), liver diseases (e.g. liver			
CC	cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),			

Db 1007 GAACGTG---TATTTCCGATGTTGAGAGATGCGGGGAACTAAGCAGAGAGATATGT 1063
Qy 1062 GGGCCCTTTCTGTGACATCTGTGAGAGACAGCTGACAAACAGCATCAGGAAACAGTGA 1121
Db 1064 GGGACCTTCTGTGAAACTGTGAGAGACAGCTTACAAACAGCATCAGAAACAGTGA 1123
Qy 1122 GAATTTGAAAGTCATGTAGAAAACCTGATACCAAGCTTCAAGGTACACTCAGAG 1181
Db 1124 GAATTTTGAAGTCATGTAGAAAAGCTTGATAACCAAGTTTCAAGGTACACTCGAAG 1183
Qy 1182 GCCTTAATGAGGAGCAGCTCATCTCAGCACAAGACCATGAGCAAGAGACTGAGAAAACCTGG 1241
Db 1184 GTTTTAATGATGAAGTGAATCTCAACGACAAACCATGAACAGAACTGGAANAACCTG 1243
Qy 1242 AGCTGAGATTTGACGGGTGTAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGAGC 1301
Db 1244 AGTTAGAAATTCAGCAGGTGTAAGAAATGATTAAGAACTCAGCAACAGCTTTTACAGCAGC 1303
Qy 1302 AGCTGGCCACCACGTGTGATGATGACACCACTCACTGTTGCGAGACTGTACTTGTCTGG 1361
Db 1304 AGCTCGCTACTGCATATGATGATGATACCACTTCCATATTACGAGACTGTTATTGTGTGG 1363
Qy 1362 AAGAAAAGGAACGCTTTAAGAAAGAGTGAGCCCTTTTAAAGAGCAAAAAGAAATTTTG 1421
Db 1364 AAGAAAAGGAACGCTCTCAAGAGAGATGTTCCCTTTTAAAGAGCAGAAAAGAAATTTTG 1423
Qy 1422 AGAGAGAAAGCGAGCTTTACAGAACTGCGCATTCGATTTGGGTTGGAGAGAAAGCGGT 1481
Db 1424 AGAGGAGAGACGAAGCTTTACAGAAAGCGCTATTCCGCTGGGATTGGAGAGAAAGCAT 1483
Qy 1482 TTGAAGAGAGCGAGCGAGCTGGCTTAAGCAGCAGTTTTAAACATCAGCACTTTCACC 1541
Db 1484 TTGAAGAGAAAGAGCGAGTGGTTAAAGAGCAGTGTCTTAAATATGACTACCTTTGACC 1543
Qy 1542 ACCAGAACTCAGAAAATGTTGAACTTTTTCAGTGCCTTCTCAGGAAGTCTTGATCCAGACA 1601
Db 1544 ACCAGAACTCAGAAAATGTTGAACTTTTTCAGTGCCTTCTCAGGAAGTCTTGATTTGGACA 1603
Qy 1602 ATCTTATAGTCCATCTCAGGCCACGGCAAGAAAGAGCTACACAGTGTGGCTAATGGGTGC 1661
Db 1604 ATCTTATAGTCACTCGAGGCGAGCGCAAGAAAGAGCTCACAGTGTGTCTAAATGGGTCTC 1663
Qy 1662 CAGCTTGACATCAAACTGATTAATCTCTTCCGCTCAGCTTCTTCTCAGCTTCTCAGCTTTC 1721
Db 1664 CAGTTTGATGCTTAAACTTAACTTAACTTCTTCTCTGCTTCCACTTCCACTTTCAGACTTTT 1723
Qy 1722 GCCAGACATTCATGTGTGCTGAACACAGTTCATCAGTGTGCTGAAATATACTCCTG 1781
Db 1724 GCCAGACACGTTCTCGCATATCTGACATAGTTCATCAATCAATGACTGAATATACTGCTG 1783
Qy 1782 AAGAAAGTAAACCAAGTGAGTTGCAAGAGAAAGCAGGATCAGAAAGTGAGAGCGTCAGT 1841
Db 1784 AAGAAATTAACCAAAATCAGGTTGGAGAGAAAGTACAAATCAAAAATGGAGTGTGGCGT 1843
Qy 1842 CGAGGCCAGCTCGGGGAGGGTGTACAGCGGATGCTCCTCGGCTTCAGGAGCGCTC 1901
Db 1844 CAAGACCTGGATACAGGAAGGTTGCTATAGTGTGCTCCTTGAGCTTACAAATATCTC 1903
Qy 1902 ACGGGACCGAGATGACTTACTTAAATGTGCGGGCTGCGTGTGTTTCCAGATGTGCG 1961
Db 1904 ATGTAGAAAAGATGACTTACTTATAGACATGTGAGCTGGAATTTTTTTCATTAATGTGT 1963
Qy 1962 CTAGAGG-----AGTTGACACAGGGGTGATGATTAAGTCAGTCGTC-----TAA 2005
Db 1964 CATCAAGTTTTCACATCTCAAGTTGAAACAGGGGTGTGTATTAAGTCAGTTATCTCTAATAA 2023
Qy 2006 CTTAAGATGCTCAGAGTTGTTGTTGAGCTTCCGCTTCCCTCCCAAGAGCTGGAATG 2065
Db 2024 CTTAAGATGCTCAGAGTTGTTGTTGAGCTTCCGCTTCCCTCCCAAGAGTTGGAATC 2083
Qy 2066 CT-AAGCTTACTTAAAGAGATCAAAAGCTTTTGG-----TTGTGTGTGTAGTAAACAGAGGCC 2120

Db 2084 TTAATCTCTATTTAAAGGATATAAAGCTTTGGATATGTATTTTAGTAACAGAGCATC 2143
Qy 2121 TGCTCTGTGACTGCAGGATGCAT-GGCGTTTGGATGGAACAGAGAGCGCTGGAATGA- 2178
Db 2144 TGGTTCGTGAAATAAAGGAATGTATAGATGTTTGGATGGAACCAAAAGCACTAGACTGAG 2203
Qy 2179 -TTGCTCGCCAGGTACCGAGAGAGACACTTTTAGGACCTGGTTCCTGTGTAACATTAAT 2237
Db 2204 TTTCTCTTATAGATATAAATATAGCCTTTTAGGAACTGATTTATGTAATGTTTAA 2263
Qy 2238 ATTCTGCCCAAGTGTGTTGGCATTGGAAGTGTAGCCCTTTACTTGAATGTATACCTGAGA 2297
Db 2264 TTTGTCTCAAAATATAGTTGGCATTGGAAGCTTTAGCCTTTACTTGAATGTATACCTGAGA 2323
Qy 2298 TTTTAAACAAGCAGGTCTCTATATTATTTATTTAGTGTGATTTTGGGATTACCTCTTT 2357
Db 2324 TTTTAAACAAGCAGGTCTCTATATTATTTATTTAGTGTG-GTTTGAATTTACCTCTTT 2382
Qy 2358 CATATGTTTT-----GTGCTGTACATAAATATACATGACTATG 2396
Db 2383 CATATGTTTTAAATAAAGTGAAATTTATGTATGTTTGTACATAGATACATGATTTATG 2442
Qy 2397 TTAAGAGGCTTTAAGGTTTAAAACTTCACACCATGCTTCAGTATAGCATTTTCATGCCAA 2456
Db 2443 TTAAGAGGCTTTAAGATTTTAAAGTTTCACA-CAACCATTAAGTATAGTATTTTCATGCC-A 2500
Qy 2457 TTAATATGTTTTCAGTGGCATGTTTACAGAGG--TTAGGACCACTGCCACATGACAG 2514
Db 2501 GTAAAAATTTTTTAGTGGTATCTGTTTACAGATGTTTAGGACCATTTGATGCAATTTACAT 2560
Qy 2515 TTAAGACTTTATTTTAAAGCCATCTGGGCAATAAATAATTCAAAAGCCCTTTCAT-AAAGCTG 2573
Db 2561 TTAAGA-ATTCTCTTTAATACATCTGGGCAATAAATATTCAGAGGTATTTCCATGAAGCTG 2619
Qy 2574 AGTTC---AGATAACTAGAACTACTAACGTTACATTTTTCAGATTTTAAAGCAATT-GTA 2629
Db 2620 AGTCTTTTAGATAATCAACACTACTAACATTTACATTTTTCAGATTTTATGACATTAGAT 2679
Qy 2630 TTTTATTTTATATATGT-GAATGTTTATATTTCTTAAGAGGAATTTTATGAGGTAAT 2688
Db 2680 TTTTATTTTGTATGTAGAAATATTATTAATTTTAAAGGACTATTGATGATAGAGAAT 2739
Qy 2689 GGGG 2692
Db 2740 AGGG 2743

RESULT 4
AAH14625
ID AAH14625 standard; cdna; 2716 BP.
XX
AC AAH14625;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:12264.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS Homo sapiens.
XX
PN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX

PA	(HELI-) HELIX RES INST.
XX	Ota T, Ieogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	WPI; 2001-318749/34.
DR	
XX	Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT	length cDNAs defined in the specification, and for the detection and/or
PT	diagnosis of the abnormality of the proteins encoded by the full-length
PT	cDNAs.
XX	
PS	Claim 8; SEQ ID NO 12264; 2537pp + Sequence Listing; English.
XX	
CC	The present invention describes primer sets for synthesising 5602 full-
CC	length cDNAs defined in the specification. Where a primer set comprises:
CC	(a) an oligo-dT primer and an oligonucleotide complementary to the
CC	complementary strand of a polynucleotide which comprises one of the 5602
CC	nucleotide sequences defined in the specification, where the
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC	of an oligonucleotide comprising a sequence complementary to the
CC	complementary strand of a polynucleotide which comprises a 5'-end
CC	sequence and an oligonucleotide comprising a sequence complementary to a
CC	polynucleotide which comprises a 3'-end sequence, where the
CC	oligonucleotide comprises at least 15 nucleotides and the combination of
CC	the 5'-end sequence/3'-end sequence is selected from those defined in the
CC	specification. The primer sets can be used in antisense therapy and in
CC	gene therapy. The primers are useful for synthesising polynucleotides,
CC	particularly full-length cDNAs. The primers are also useful for the
CC	detection and/or diagnosis of the abnormality of the proteins encoded by
CC	the full-length cDNAs. The primers allow obtaining of the full-length
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB95893 to
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC	oligonucleotides, all of which are used in the exemplification of the
CC	present invention
XX	
SQ	Sequence 2716 BP; 935 A; 405 C; 524 G; 852 T; 0 U; 0 Other;
	Query Match 46.8%; Score 1260.4; DB 4; Length 2716;
	Best Local Similarity 79.7%; Pred. No. 0;
	Matches 1716; Conservative 0; Mismatches 376; Indels 60; Gaps 17;
Qy	594 TCCTGAAGAATGAGAAAGATCAGGTACAAAAATTACAAAATATCATAGCCGGCGCTA 653
Dd	
	1 TACTAAAGAATGAGAAAGATCAGGTGC AAAAATTACAAAATATCATTCGAGTCGAGCTA 60
Qy	654 CTCAGTATAATCATGTGTGAAGAGGAGGAGCGTGTAATATAATAGCTAAAGGAGGCC 713
Dd	
	61 CTCAGTATAATCATGTATGAAGAGAGAAAGCGGTGAATATAATAAATGAAGGAAGCTC 120
Qy	714 TGCATCAGCTCGTTATGAAACAAGAGGATAAAACATAGCCATCGATGTTTTTAATTTATG 773
Dd	
	121 TACATCAACTTGTTATGAACAAGAAAGATAAGAAAATAGCTATGGAATTTTGAATTTATG 180
Qy	774 TGGGTCGAGCTGATGGCAAACCGAGGCTCATGGAGGACTGCAGAAACAGAGCCAGGAATG 833
Dd	
	181 TCGGGAGAGCTGATGGAAAAGAGAGCTCTCGAGGACTGTTGAANAATGAGCCAGGAATG 240
Qy	834 AAGATGAGATGTA CAATAATTCTGTTGAATGATTAATAGTAGTCCGCCAGAACAGATCTCTGA 893
Dd	
	241 AAGATGAAATGTTATAAAAATTTCTCTTTGAATGATTATGAATATCGTCAAGAAACAAATCTCTAA 300
Qy	894 TGGAGAACGGGAGCTCAGAAAGAGTCTCTCCAGCAGATGMAAGAGGATGATCTCTCTCCC 953
Dd	
	301 TGGAAATGCAAACTTAAGAAAGTCTTTCACCAAAATGAAAAGGAATGATTTCTCTTC 360
Qy	954 TGTCCTCTCAGAAAGAACGCCAGGGAAAGAGCAGAGGACGGCACAGGCACTCTTGCTA 1013
Dd	
	361 TTCTCTCCCCAAAGAAACCTAGAGAAAGATAGATGATGATGACGAGGAATCTG---TA 417
Qy	1014 TCTCCGATATAGAAAGATGACTCTCTGGGAACTGAGCAGAGACAGCGTGTGGGGCTCTTTCCT 1073

PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
XX Claim 1; SEQ ID NO 3761; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
XX Sequence 826 BP; 329 A; 120 C; 179 G; 195 T; 0 U; 3 Other;
SQ
Query Match 20.8%; Score 560; DB 4; Length 826;
Best Local Similarity 82.1%; Pred. No. 3.3e-142;
Matches 680; Conservative 0; Mismatches 142; Indels 6; Gaps 3;
594 TCCTGAAGATGAGAAGATGAGGTACAAAATATACAAAATATCATAGCCAGCGGGCTA 653
Db 1 TACTAAAGATGAGAAGATGAGGTGCAAAATATACAAAATATCATTCGACAGTCGAGCTA 60
654 CTCAGTATATCATGATGTGAAGAGGAGGAGCGTGAATATATTAAGCTAAAGAGCGCC 713
Db 61 CTCAGTATATCATGATGTGAAGAGGAGGAGCGTGAATATATTAAGCTAAAGAGCGTC 120
714 TGCATCAGCTGTTATGAACAGAGGATGAACATGATGATGATGATGATGATGATGATGATG 773
Db 121 TACATCAACTGTTATGAACAGAGGATGAACATGATGATGATGATGATGATGATGATGATG 180
774 TGGGTGCGAGCTGATGGCAAGAGGCTCATGGAGGAGCTGACAAAACAGAGCCAGGAGATG 833
Db 181 TCGGAGAGCTGATGGAAAGAGGCTCTCGAGGAGCTGGTAAACTGAAGCCAGGAGATG 240
834 AAGATGAGATGTACAAAATCTGTTGAATGATGATGATGATGATGATGATGATGATGATGATG 893
Db 241 AAGATGAAATGTATAAAATCTCTTGAATGATGATGATGATGATGATGATGATGATGATGATG 300
894 TGGAGACGCGAGCTGAAGAGGTCCTCCAGCAGATGAAGAAGAGAGATGATCTCTCC 953
Db 301 TGGAAATGCAAGACTTGAAGAGGTCCTCCAGCAGATGAAGAAGAGAGATGATCTCTCC 360
954 TGTCTCTCTCAGAAGAGAGCCCGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1013
Db 361 TTTCTCCCAAAAGAGAGAACTAGAGAAAGAGTAGATGATGATGATGATGATGATGATGATGATG 417
1014 TCTCGATATAGAAGATGACTCTGGGAACTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1073
Db 418 TTTCCGATGTTGAAGAGATGCGGGAACTAAGCAGAGAGAGATGATGAGGACCTTTCTCT 477

QY 1074 GTGACACTGTGAGAGAGCAGCTGACAAACAGCATCAGGAAACAGTGGAGAAATTTTGA AAA 1133
Db 478 GTGAACTGTGAGAGAGCAGCTTACAAACAGCATCAGAAACAGTGGAGAAATTTTGA AAA 537
QY 1134 GTCATGTAGAAAACTCGATAACCAAGCTTCGAAGGTACACTCGAGAGCGCTTAAATGAGG 1193
Db 538 GTCATGTAGAAAAAGCTTGATAACCAAGCTTCAAAGGTACACCTCGAAAGGTTTAAATGATG 597
QY 1194 AGNAGCTCATCTCAGCACAAGACCATGAGCAGAGAGACTGAGAACTCGAGCTGAGATTG 1253
Db 598 AAGATGTAATCTCAGCAACAGACCATGAACAAGAAACTGAAAAAACTCGAGTTGAAAAATTC 657
QY 1254 AGCGGTGTAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCA 1313
Db 658 AGCAGTGAAGAAATGATTAAGCTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTG 717
QY 1314 CGTGTGATGATGACACCACT 1372
Db 718 CATATGATGATG--ATCCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 775
QY 1373 CGCCTTAAAGAGAGTGGACCTTTTAAAGAGCAAAAAGAAATTTT 1420
Db 776 CGTNTCAAGAAAAAATGGGCCCTTTTAAAGAAACCCNAAAAAGAAAT 823
RESULT 7
ADQ78551
ID ADQ78551 standard; DNA; 668 BP.
XX
AC ADQ78551;
XX
DT 21-OCT-2004 (first entry)
XX
DE Novel canine microarray-related DNA sequence SeqID10110.
XX
KW canine microarray; drug screening; toxicity assay;
KW environmental pollutant; cellular response; gene expression profile;
KW toxic response; liver necrosis; fatty liver disease;
KW protein adduct formation; hepatitis; dog; ds.
XX
OS Canis familiaris.
XX
PN WO2004063324-A2.
XX
PD 29-JUL-2004.
XX
PF 05-MAY-2003; 2003WO-US013853.
XX
PR 03-MAY-2002; 2002US-0377240P.
XX
PA (GENE-) GENE LOGIC INC.
PA (PFIZ) PFIZER PROD INC.
XX
PI Diggins JC, Porter M, Wei T;
XX WPI; 2004-561890/54.
XX
XX New isolated nucleic acid molecule, useful for drug screening and
PT toxicity assays or for assessing the impact, including toxicity, of a
PT compound, pharmaceutical agent or environmental pollutant on a cell or
PT living organism.
XX
PS Claim 1; SEQ ID NO 10110; 41pp; English.
XX
CC This invention is related to a novel isolated canine nucleic acid
CC sequences and the construction of canine microarrays containing a
CC significant portion of the canine genome. The isolated canine nucleic
CC acid sequences of the invention may be useful for drug screening and
CC toxicity assays. The invention is therefore useful for assessing the
CC impact, including toxicity, of a compound, pharmaceutical agent or
CC environmental pollutant on a cell or living organism. The methods are
CC useful for detecting genes that are up- or down-regulated in canines in a
CC disease state. The sequences are useful as diagnostic agents or markers


```
QY 2033 GACTTCGCTGTCTTCCGCCAAAGAGCTGAAATGCT-AAGCTACTTTAAAGGATGCAAGC 2091
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3216 GACTTCCCTGTCTTCCGCCAAAGAGTTGAAATCTTAATCTATTTAAAGGATATAAAG 3275
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2092 TTTCG---TTGTGTCTTAGTAACAGAAGCCCTGCTCTGTGACTGCAGGAATGCAT-G 2146
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3276 CTTTGGATATGATTTTATTAGTAACAGAGCACTGCTGTTCTGTGAATAAAGGAATGATAG 3335
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2147 GCGTTTGGATGGAACAGAGAAGCGCTGGAATGA--TTGCTCCGCCAGGTACCCGAGAAGC 2204
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3336 ATGTTTGGATGGAACAAAGACACTAGACTGAGTTTCTCTTATAGGTATTAATAATAGC 3395
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2205 ACTTTTAGGACTGTGTTCTGTTAAACATTAATATTCGTCCTCCAAAGTGTGGTGGCATGG 2264
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3396 ACTTTTAGGAAACTGATTAATGTTAAATGTTTAAATTTGTTCTCAAAATATAGTTGGCATGG 3455
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2265 AAGTGTAGCCTTTACTTGAATGTATCTGTAGATTTTAAACAAAGCAGGTTCTATATTTA 2324
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3456 AAGTTTAGCCTTTACTTGAATGTATCTGTAGATTTTAAACAAAGCAGGTTCTATATTTA 3515
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2325 TTATGTTAGTGTGATTTTGGGATACCTCTTTTCATATGTTT----- 2367
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3516 TTATGTTTAGTGTG-GTTTGAAATTAACCTCTTTTCATATGTTTAAATAAGTGAATTTA 3574
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2368 ----GTGCTGTACATAATATACATGATGTTTAAAGGCTTTAAAGTTTAAAACTT 2423
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3575 TGATGTTTGTACATAGATACACATGATTAATGTTTAAAGAGGCTTTAAAGATTTAAAAAGTTT 3634
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2424 CACACATGCTTGAGTATAGCATTTTCATGCCAATTAATAATGTTTTCAGTGGCATGTTT 2483
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3635 CACA-CAACCATAGTATAGTATTTTCATGCC-AGTAAATTTTTTATGTTGGTATTTCTGTT 3692
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2484 TACAGAGG--TTAGGACCACTCCACATGACAGTTAAGACTTTTATTTTAAAGCCATCTGG 2541
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3693 TACAGATGATTAGGACCATTCATGCAATTTACATTTAAGA-ATTCCTTTAATACATCTGG 3751
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2542 GCNATAAAANTCAAAGCCCTTCAT-AGCTGAGTTC---AGATACTAGAACTACTAA 2597
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3752 GCAATTAATATTGAAAGGATTTCCCATGAGCTGAGTCTCTTTAGATAATCAACACTACTAA 3811
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2598 CTTTACATTTTGTAGATTTTAAAGCATT-GTATTTTATTTTATATATGT-GAATGTTAT 2655
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3812 CATACATTTTGTAGATTTTATGACATAGATTTTATTTTGTATATGATAGATATTAT 3871
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2656 AATTTCTAAGAGGAATTTGATTAATGGAGTAATGGG 2692
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3872 AATTTTAAAGGACTATTGATGATAGAGTAATAGG 3908
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
ABA20638
ID ABA20638 standard; DNA; 6619 BP.
XX AC ABA20638;
XX XX
XX DT 23-JAN-2002 (first entry)
XX DE Human nervous system related polynucleotide SEQ ID NO 12969.
XX KW Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX OS Homo sapiens.
XX PN WO200159063-A2.
XX PD 16-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001334.
XX PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 25-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 26-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
```


QY	2368	---GTCTGTGTACATAAATATACATGACTATGTTAAAGAGGCTTTAAAGCTTTAAAAACTT	2423	PR	14-AUG-2000;	2000US-0225214P.			
Db	3575	TGTATGTTTGTACATAGATACACATGATGTTAAAGAGGCTTTAAAGATTTAAAAAGTTT	3634	PR	14-AUG-2000;	2000US-0225266P.			
QY	2424	CACACCATGCTTGAGTATAGCAATTTTCATGCCAATTAATAATGTTTTCAGTGGCATGTGTT	2483	PR	14-AUG-2000;	2000US-0225270P.			
Db	3635	CACA-CAACCAATAAGTATAGTATTTTCATGCC-AGTAAATTTTAAAGTGGTATCTGTT	3692	PR	14-AUG-2000;	2000US-0225447P.			
QY	2484	TACAG--AGTTAGGACCACTGCCACATGACAGTTAAAGACTTTATTTTAAAGCCATCTGG	2541	PR	14-AUG-2000;	2000US-0225757P.			
Db	3693	TACAGATATATTAGGACCACTTCATGCAATTCATACATTTAAGA-ATTCTCTTTAATACATCTGG	3751	PR	14-AUG-2000;	2000US-0225758P.			
QY	2542	GCAATAAAANTCAAAGCCCTTCAT-AGCTGAGTTC---AGATACTAGAACTACTAA	2597	PR	18-AUG-2000;	2000US-0225759P.			
Db	3752	GCAATAAATATTGAAGGTAATTCATGAAGCTGAGTTCCTTTAGATAATCAACACTACTAA	3811	PR	22-AUG-2000;	2000US-0226681P.			
QY	2598	CGTTACATTTTTCAGATTTTAAAGCAAT-GTATTTTATTTTATATATGT-GAATGTTAT	2655	PR	22-AUG-2000;	2000US-0226868P.			
Db	3812	CATTACATTTTTCAGATTTTATGACATTAGATTTTATTTTGTATATGTAGAAATATTAT	3871	PR	23-AUG-2000;	2000US-0227009P.			
QY	2656	AATTTCTAGAGGAATATTGATTATGGAGTAATGGGG	2692	PR	30-AUG-2000;	2000US-0228924P.			
Db	3872	AATTTTAAAGGACTATTGTATGATAGAGAATAGGG	3908	PR	01-SEP-2000;	2000US-0229287P.			
RESULT 10						PR	01-SEP-2000;	2000US-0229343P.	
ID	ABA20637	standard; DNA; 6619 BP.					PR	01-SEP-2000;	2000US-0229344P.
XX	XX						PR	05-SEP-2000;	2000US-0229345P.
AC	ABA20637;						PR	05-SEP-2000;	2000US-0229509P.
XX	XX						PR	06-SEP-2000;	2000US-0229513P.
DT	23-JAN-2002.	(first entry)					PR	06-SEP-2000;	2000US-0230437P.
XX	XX						PR	08-SEP-2000;	2000US-0230438P.
DE	Human nervous system related polynucleotide SEQ ID NO 12968.						PR	08-SEP-2000;	2000US-0231242P.
KW	Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;						PR	08-SEP-2000;	2000US-0231243P.
KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;						PR	08-SEP-2000;	2000US-0231244P.
KW	antiparkinsonian; antiscikling; antianaemic; antiarthritis; cancer;						PR	08-SEP-2000;	2000US-0231413P.
KW	antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;						PR	08-SEP-2000;	2000US-0232080P.
KW	antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;						PR	08-SEP-2000;	2000US-0232081P.
KW	antiparasitic; cardiant; immune disorder; cardiovascular disorder;						PR	12-SEP-2000;	2000US-0231968P.
KW	neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.						PR	14-SEP-2000;	2000US-0232397P.
OS	Homo sapiens.						PR	14-SEP-2000;	2000US-0232398P.
XX	XX						PR	14-SEP-2000;	2000US-0232400P.
PN	WO200159063-A2.						PR	14-SEP-2000;	2000US-0232401P.
XX	XX						PR	14-SEP-2000;	2000US-0233063P.
XX	16-AUG-2001.						PR	14-SEP-2000;	2000US-0233064P.
PF	17-JAN-2001; 2001WO-US001334.						PR	14-SEP-2000;	2000US-0233065P.
XX	31-JAN-2000; 2000US-0179065P.						PR	21-SEP-2000;	2000US-0234223P.
PR	04-FEB-2000; 2000US-0180628P.						PR	25-SEP-2000;	2000US-0234274P.
PR	24-FEB-2000; 2000US-0184664P.						PR	25-SEP-2000;	2000US-0234987P.
PR	02-MAR-2000; 2000US-0186350P.						PR	26-SEP-2000;	2000US-0234998P.
PR	16-MAR-2000; 2000US-0189874P.						PR	26-SEP-2000;	2000US-0235484P.
PR	17-MAR-2000; 2000US-0190076P.						PR	27-SEP-2000;	2000US-0235834P.
PR	18-APR-2000; 2000US-0198123P.						PR	29-SEP-2000;	2000US-0236327P.
PR	19-MAY-2000; 2000US-0205515P.						PR	29-SEP-2000;	2000US-0236367P.
PR	07-JUN-2000; 2000US-0209467P.						PR	29-SEP-2000;	2000US-0236368P.
PR	28-JUN-2000; 2000US-0214886P.						PR	29-SEP-2000;	2000US-0236369P.
PR	30-JUN-2000; 2000US-0215135P.						PR	02-OCT-2000;	2000US-0237037P.
PR	07-JUL-2000; 2000US-0216647P.						PR	02-OCT-2000;	2000US-0237039P.
PR	07-JUL-2000; 2000US-0216880P.						PR	02-OCT-2000;	2000US-0237040P.
PR	11-JUL-2000; 2000US-0217487P.						PR	13-OCT-2000;	2000US-0239935P.
PR	11-JUL-2000; 2000US-0217496P.						PR	20-OCT-2000;	2000US-0239937P.
PR	14-JUL-2000; 2000US-0218290P.						PR	20-OCT-2000;	2000US-0240960P.
PR	26-JUL-2000; 2000US-0220963P.						PR	20-OCT-2000;	2000US-0241785P.
PR	26-JUL-2000; 2000US-0220964P.						PR	20-OCT-2000;	2000US-0241786P.
PR	14-AUG-2000; 2000US-0224518P.						PR	20-OCT-2000;	2000US-0241808P.
PR	14-AUG-2000; 2000US-0224519P.						PR	20-OCT-2000;	2000US-0241809P.
PR	14-AUG-2000; 2000US-0225213P.						PR	20-OCT-2000;	2000US-0241826P.
PR							PR	20-OCT-2000;	2000US-0242221P.
PR							PR	01-NOV-2000;	2000US-0244617P.
PR							PR	08-NOV-2000;	2000US-0246474P.
PR							PR	08-NOV-2000;	2000US-0246475P.
PR							PR	08-NOV-2000;	2000US-0246476P.
PR							PR	08-NOV-2000;	2000US-0246477P.
PR							PR	08-NOV-2000;	2000US-0246523P.

PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system cancers
PT and metastases.
XX
XX Disclosure; SEQ ID NO 12969; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABA11004-ABA215134) and proteins
CC (ABR14678-ABR18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing
CC; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 6619 BP; 2137 A; 1013 C; 1233 G; 2236 T; 0 U; 0 Other;
Query Match 14.9%; Score 399.8; DB 5; Length 6619;
Best Local Similarity 74.0%; Pred. No. 8.2e-98;
Matches 738; Conservative 0; Mismatches 202; Indels 57; Gaps 16;
Qy 1749 ACAAGTTCCATCAGTGTGCTGAATATATACTCCTCGAAGAAAGTAAACCAAGTGAAGGTTGCAA 1808
Db 2916 ACAGTTCAATCAATGTAATATACTGCTGAGAAATTAACCAAAATCAGGTTGGAG 2975
Qy 1809 GAGAAAGCAGGATCAGAAAGTCAGTCGAGTCGAGCCGAGCTCCGGGAGGGGTGCT 1868
Db 2976 GAGAAATGTACAAATCAAAATGAGTGTGGCATCAAGACCTGGATCAAGAGAGGTTGCT 3035
Qy 1869 ACAGCGATGCTCCTCGGCTTCAGGAGCGCTCAGGGGACCGAGATGACTTACCTTAAA 1928
Db 3036 ATAGTGGATGCTCCTTGGAGCTACACAAATTTCTCATGTAGAAAAGATGACTTACCTTAGA 3095
Qy 1929 TGTGGGGCTGCAGTGTCTCCAGATGTCGCGCTAGAGG-----AGTTGACAC 1977
Db 3096 CATGTGGACTGGAATTTTTTTCATTAAATGTTTCATCAAGTTTCACATCTAAGTTGAAAC 3155
Qy 1978 AGGGTGTAGCATAAAGTCAGTCGTC-----TAACTTAAAGATGTCAGAGTTGTTTGTG 2032
Db 3156 AGGGTGTGTATAAAGTCAGTTATCTCTAATAACTTAAAGATGGTCTGAGTTGTTGTTG 3215
Qy 2033 GACTTCGCTGCTCTCCCGGAAAGCTGAAATGCT-NAGCTACTTAAAGGATGCAAGC 2091
Db 3216 GACTTCCTGCTCTCCCGGAAAGGTTGAAATCTTAAATCTATTTAAAGGATATAAAG 3275
Qy 2092 TTTGG----TTGTGTGTAGTAACAGAAAGCCCTGGCTCTGTGACTCGAGGAATGCAT-G 2146
Db 3276 CTTTGGATATGATTTTATAGTACAGAAAGCATCTGGTCTCTGTAATAAGGAATGTATAG 3335
Qy 2147 GCGTTTGGATGGAACAGAGCGCTGGAATGA--TTGCCCTCGCAGGTACCGAGAGAGC 2204
Db 3336 ATGTTTGGATGGAACACAAAGACACTAGACTGAGTTTCCTCTATATAGGTATTAATAATAGC 3395
Qy 2205 ACTTTAGGAGCTGGTTCCTGTAAACATTAATAATTCGTCCCAAGCTGGTTGGCATTGG 2264
Db 3396 ACTTTAGGAAACTGATATTGTAATAATTTGTTAAATTTGCTCTCAATAATAGTTGGCATTGG 3455
Qy 2265 AAGTGTAGCCTTTACTTTGAATGTACTGTAGATTTTAAACAAAGCAGGTCTTATATTTA 2324
Db 3456 AAGTTAGCCTTTACTTTGAATGTACTGTAGATTTTAAACAAAGCAGGTCTTATATTTA 3515
Qy 2325 TTATGTTTGTGATTTTGGGATTAACCTCTTTTCATATGTTTT----- 2367
Db 3516 TTATGTTTGTGATG-TTTTGAATTTACCTCTTTTCATATGTTTTTAAATFAAAGTGAATTTA 3574
Qy 2368 ----GTGCTGTACATAAATATACATGACTATGTTAAAGAGCTTTAAAGGTTTAAAGACTT 2423
Db 3575 TGTATGTTTGTACATAGATACACATGATTTATGTTAAGAGGCTTTAAGATTTAAGAGTTT 3634
Qy 2424 CACACCATGCTTGGATATAGCATTTTCATGCAATTTAAATTTTTCAGTGGCATGGTGT 2483
Db 3635 CACA-CAACCATAGTATAGTATTTTCATGCC-AGTAAATTTTATAGTGGTATCTGTT 3692
Qy 2484 TACAGAG--TTAGGACCACTGCCCATGACAGTTAAGACTTTATTTTAAAGCCATCTGG 2541
Db 3693 TACAGATGTTATTAGGACCACTGATGCAATTTAAGATTTTAAAGATTTTAAAGTTT 3751
Qy 2542 GCATTAATAATTCAGAGCCCTTTCAT--AAGCTAGTTTC--AGATAACTAGAACTACTAA 2597
Db 3752 GCATTAATAATTCAGAGGATTTTCCATGAGCTGAGTTCTTTAGATTAATCAACACTACTAA 3811
Qy 2598 CGTTACATTTTTCAGATTTTAAAGCAAT--GTATTTTATTTTATATATGT--GAATGTAT 2655
Db 3812 CATTACATTTTTCAGATTTTATGACATTAGATTTTATTTTATTTTGTATATGTAGATATAT 3871
Qy 2656 AATTTCTAAGAGGATTTTATGATTTATGAGGTAATGGGG 2692
Db 3872 AATTTTAAAGGAGCTATTGATGATAGAAGATAGGG 3908

Claim 3; SEQ ID NO 5746; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03366 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH95893 to AAH13632 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 563 BP: 144 A: 115 C: 97 G: 199 T: 0 U: 8 Other:

Query Match 12.4%; Score 334.2; DB 4; Length 563;
Best Local Similarity 85.7%; Pred. No. 1.8e-80;
Matches 383; Conservative 0; Mismatches 63; Indels 1

288	ATCAGAGCTGACCAACCTTCGGGTTTCCTTCCTTGATGGAAGATCCAAAGTAAAGAGG	347
Qy		
447	ACCAGGAATTGACTACTTTTGGTTTCCTTCATTATATGAAGAAATCCAAAGTAAAGAGA	388
Db		
348	CAAAAGAGAGAAATTAATATAGTCGCTGTTCTGAACTGTATGAACAGAGCTGCTCGTCTTC	407
Qy		
387	CAAAAGAGAGGTTAAATATAGTAGCTGTACTAAATTTGTATGAATGAGCTGCTTGCTCTTC	328
Db		
408	AGCGGAAGAACCTGCTGCGCCCAAGGAGAGCGTGGAGACACAGAACTTGAAGCTCGGCAGTG	467
Qy		
327	AGCGGAAGAACCTTCTAGCTCAGGAAAAATGTGGAGACACAGAAATTTGAAGCTGGGAAGTG	268
Db		
468	ACATGAGCACCACTGACAGAGCTCTACGCCAACTTTAAGAGGCAGTTGGGAAACGTCACGCG	527
Qy		
267	ATATGGACCATCTACAGAGCTGCTACTCAAAACCTTTAAGGAACAACTGGAAACCTCCAGGA	208
Db		
528	GGGAGATGATCGGGCTTCAAGAGAGAGACAGCGCAGCTGCAGTGCAGAAACAGGAGTTTGC	587
Qy		
207	GGGAAATGATTGGGCTTCAGGAAAGAGACACAGCAGTTACAATGTGAAGAACAGGAAATTCG	148
Db		
588	ATCAGCTCCTGAAGAATGAGAAAGATGAGGTAC-AAAAATTCNAAATATCATAGCCAGC	646
Qy		
147	ATCAGCTACTAAAGAATGAGAAAGATGAGGTGCAAAAANATTCACAAATATCATTTGCAAGT	88
Db		
647	CGGGCTACTCAGTATAATCATGATGTGAAGAGGAGGAGCGTGAAATATAATAAGCTAAAG	706
Qy		
87	CGAGCTACTCAGTATATCATGATATGAAGAGAGAAAGAGCGTGATATATATAACTGAAG	28
Db		
707	GAGCGCTCGCATCAGCTCGTTATGAAC	733
Qy		
27	GAACTGCTACATCAACTGGTTATGAAC	1
Db		

RESULT 13

RESULTS 13
AAV86670
ID AAV86670 standard; cDNA: 498 BP.

AAV86670;

DT 27-APR-1999 (first entry)

CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors
XX
SQ Sequence 428 BP; 170 A; 69 C; 94 G; 95 T; 0 U; 0 Other;
Query Match 11.0%; Score 297.2; DB 3; Length 428;
Best Local Similarity 82.3%; Pred. No. 2.1e-70;
Matches 354; Conservative 0; Mismatches 73; Indels 3; Gaps 1;
Qy 878 CAGAAGCAGATCCTGTAGGAGAACCGGAGCTGAAGAAGTCTCCAGCAGATGAAGAAG 937
Db 1 CAGAAACAAATCCTAATGGAATATGCAGAACTTAAGAAGGTTCTTCAACAAATGAAAAAG 60
Qy 938 GAGATGATCTCTCTCTCTCTCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 997
Db 61 GAAATGATTTCTCTCTCTCTCTCTCTCCCAAGAGAGAGAGAGAGAGAGAGAGATGAT 120
Qy 998 ACAGCACTGTGTCTATCTCCGATATAGAGATGACTCTGGGGAACTGAGCAGAGACAGC 1057
Db 121 ACAGGAAGCTGT---TATTTCCGATGTTGAAGAGATGCCGGGAACTAAGCAGAGAGT 177
Qy 1058 GTGTGGGGCCCTTCTCTGTGACACTGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
Db 178 ATGTGGGACCTTCTCTGTGAACTGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAG 237
Qy 1118 TGGAGAAATTTGAAAGTCATGTAGAAAACCTCGATAACCAAGCTTCGAGGTACACTCA 1177
Db 238 TGGAGAAATTTGAAAGTCATGTAGAAAAGCTTGTATAACCAAGTTTCANAGGTACACTG 297
Qy 1178 GAGGCGCTTAATGAGGAGGAGTCTATCTCAGCAAGACCATGAGCAAGAGACTGAGAAA 1237
Db 298 GAAGTTTAAATGATGAAGATGTAATCTCAGCAAGAGCCATGAACAGAACTGAAGAAA 357
Qy 1238 CTGAGCTGGAGATGTAGCGGTGTAAGAGATGATCAAGGCTCAGCAGAGAGCTCTTACAG 1297
Db 358 CTCGAGTTAGAAATTCAGCAGTGTAAAGAAATGATTAAAACTCAGCAACAGCTTTTACAG 417
Qy 1298 CAGCAGCTGG 1307
Db 418 CAGCAGCTCG 427

Search completed: June 13, 2006, 20:46:44
Job time : 1470 secs

Result No.	Score	Query Match	Length	DB	ID	Description	
1	3165	100.0	615	8	ADO26491	Murine af	
2	2937	92.8	613	8	ADO26493	Rat afadi	
3	2598.5	82.1	567	7	ABO14751	Novel hum	
4	2152.5	68.0	504	7	ADB65313	Human pro	
5	1855.5	58.6	417	4	AA993320	Human pro	
6	612.5	19.4	137	3	AAG30300	Human sec	
7	379.5	12.0	483	8	ADx91485	Plant ful	
8	343	10.8	382	3	AGA40470	Arabidops	
9	310	9.8	347	3	AGA48364	Arabidops	
10	290	9.2	290	8	ADx76571	Plant ful	
11	277	8.8	291	3	AAG33165	Zea may	
12	277	8.8	307	3	AGA48365	Arabidops	
13	265.5	8.4	288	3	AGA06714	Arabidops	
14	262.5	8.3	277	3	AAG33166	Zea may	
15	258	8.2	284	3	AGA40471	Arabidops	
16	244	7.7	281	3	AGA48366	Arabidops	
17	236.5	7.5	1788	4	AAW40467	Human pol	
18	234.5	7.4	1679	8	ADs43650	Bacterial	
19	232	7.3	235	3	AAG33167	Zea may	
20	225	7.1	2503	9	ADW44357	Human cen	
21	224	7.1	712	9	AEAs1202	Chicken R	
22	223.5	7.1	3259	7	DEB56037	Human Pro	
23	223.5	7.1	3259	7	ADBS56033	Human Pro	

Query Match	100.0%	Score 3165;	DB 8;	Length 615;			
Best Local Similarity	100.0%;	Pred. No. 5.2e-215;					
Matches 615;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;			
Qy	1	MGDMMTVDPVLCCTENKNL	SOYTTSETKMS	PSLSY	SOQVLCSSVPLSKNVHGVFGVFCCTGE	60	
Db	1	MGDMMTVDPVLCCTENKNL	SOYTTSETKMS	PSLSY	SOQVLCSSVPLSKNVHGVFGVFCCTGE	60	
Qy	61	NIEQSI	SYLDQELTTFGPPS	LYESK	EAARELNIVAVLNCMNE	LLVQRLKLLAQAQSV	120
Db	61	NIEQSI	SYLDQELTTFGPPS	LYESK	EAARELNIVAVLNCMNE	LLVQRLKLLAQAQSV	120
Qy	121	ETQNLKLGSDMDH	LQSCYAKLKEO	LTSRRRMI	GLQERDRLOLCKNRSL	HLHOLLKNEKDEV	180
Db	121	ETQNLKLGSDMDH	LQSCYAKLKEO	LTSRRRMI	GLQERDRLOLCKNRSL	HLHOLLKNEKDEV	180
Qy	181	QKLQNI	ITASRATQYNH	DOVKRERYNK	LKERLHOLVMNKDKNT	AMDVLVNVGRADGKRG	240
Db	181	QKLQNI	ITASRATQYNH	DOVKRERYNK	LKERLHOLVMNKDKNT	AMDVLVNVGRADGKRG	240
Qy	241	SWRTDKTEARN	DEMYKILLND	YEYRQKQIL	MENAEELKKVIL	QMKKEMISLSPQKKKPR	300
Db	241	SWRTDKTEARN	DEMYKILLND	YEYRQKQIL	MENAEELKKVIL	QMKKEMISLSPQKKKPR	300
Qy	301	ERAEDGTGTVA	ISDIEDSGELS	RDSDVWG	LSCDTVRQOLTNSIR	KQWRILKSHVEKLDNQ	360
Db	301	ERAEDGTGTVA	ISDIEDSGELS	RDSDVWG	LSCDTVRQOLTNSIR	KQWRILKSHVEKLDNQ	360
Qy	361	ASKVHSEGLNE	EDVISQDHEQ	ETEKLLETER	CKEMIKAQOQL	LQOQLATTCDDDTTSL	420
Db	361	ASKVHSEGLNE	EDVISQDHEQ	ETEKLLETER	CKEMIKAQOQL	LQOQLATTCDDDTTSL	420
Qy	421	LRDCYLLEEK	ERLKEEWTL	FKEQKKNFER	RRSFTEAAIR	LGLERKAFEEERASWVKQF	480
Db	421	LRDCYLLEEK	ERLKEEWTL	FKEQKKNFER	RRSFTEAAIR	LGLERKAFEEERASWVKQF	480
Qy	481	LNMNTPHQNS	ENVKLP	SFAFSGSD	DPNLI	VHSPRQKLSHVANGVPACTSKLTKSLPA	540
Db	481	LNMNTPHQNS	ENVKLP	SFAFSGSD	DPNLI	VHSPRQKLSHVANGVPACTSKLTKSLPA	540
Qy	541	SPGTS	DPQTHSCVSE	HSSISVLN	ITPEESK	PSFVARESTDQKWSVQSRPSRSGCYSGC	600
Db	541	SPGTS	DPQTHSCVSE	HSSISVLN	ITPEESK	PSFVARESTDQKWSVQSRPSRSGCYSGC	600
Qy	601	SSAFRS	AHGRDRDL	P	615		
Db	601	SSAFRS	AHGRDRDL	P	615		

[illegible]

RESULT 3
ABO14751
ID ABO14751 standard; protein; 567 AA.
XX
AC ABO14751;
XX
DT 25-AUG-2003 (first entry)
XX
DE Novel human protein #124.
XX
KW Human; NOV; gene therapy; endocrine related disease; diabetes;
KW metabolism-related disease; obesity; central nervous system disorder;
KW Alzheimer's disease; Parkinson's disease; epilepsy; multiple sclerosis;
KW schizophrenia; depression; autoimmune disorder; inflammatory disorder;
KW psoriasis; allergy; lupus erythematosus; asthma; cancer;
KW inflammatory bowel disease; rheumatoid arthritis; osteoarthritis;
KW colon cancer; liver cancer; breast cancer; ovarian cancer;
KW prostate cancer; brain cancer; melanoma; liver disease; liver cirrhosis;
KW lung disease; emphysema; obstructive pulmonary disease; haemophilia;
XX stroke; infection.
XX
OS Homo sapiens.
XX
PN WO2003023002-A2.
XX
PD 20-MAR-2003.
XX
PF 09-SEP-2002; 2002WO-US028539.
XX
PR 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318130P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322836P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 26-SEP-2001; 2001US-0324990P.
PR 17-APR-2002; 2002US-0373212P.
PR 06-SEP-2002; 2002US-00236177.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Spytek KA, Patturajan M, Gorman L, Li L, Anderson DW, Zhong M;
PI Gerlach VL, Vernet CAM, Ellerman K, Berghs C, Rothenberg ME, Guo X;
PI Shinkets RA, Leach MD, Catterton E, Kekuda R, Ji W, Miller CE;
PI Rieger DK, Taupier RU, Shenoy SG, Liu X, Padigaru M, Alsobrook JP;
PI Lepley DM, Edinger SN, Burgess CE;
XX
WPI; 2003-313242/30.
DR N-PSDB; ACD19444.
XX
PT New cytoplasmic, nuclear membrane bound or secreted polypeptides (NOV)
PT and polynucleotides, useful in gene therapy, e.g. for treating or
PT preventing obesity, multiple sclerosis, allergy, cancers, hemophilia,
PT stroke or infections.
XX
PS Claim 1; Page 353; 586pp; English.
XX
CC The invention describes a new isolated polypeptide (NOV). The NOVX
CC polypeptide, nucleic acid and antibody are useful as therapeutics,
CC particularly in the manufacture of a medicament for treating a syndrome
CC associated with a human disease, which includes a pathology associated
CC with NOVX polypeptide. The DNA encoding the protein is useful in gene
CC therapy for treating the disease or condition. In particular, the NOVX
CC polypeptide or polynucleotide is useful for treating endocrine/
CC metabolism-related diseases (e.g. obesity or diabetes), central nervous
CC system disorders (e.g. Alzheimer's disease, Parkinson's disease,
CC epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune

CC and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,
CC asthma, inflammatory bowel disease, rheumatoid arthritis or
CC osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,
CC prostate or brain cancers, or melanoma), liver diseases (e.g. liver
CC cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),
CC haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).
CC These are also useful in developing powerful assay system for functional
CC analysis of various human disorders, as well as in diagnostic
CC applications, and for monitoring the effects of drugs during clinical
CC trials. This is the amino acid sequence of a novel human NOV protein
XX
SQ Sequence 567 AA;
Query Match 82.1%; Score 2598.5; DB 6; Length 567;
Best Local Similarity 89.1%; Pred. No. 5.5e-175;
Matches 506; Conservative 26; Mismatches 35; Indels 1; Gaps 1;
QY 28 MPSSLYSOOVLCSYVPLSKNVHGVGVCTGNIQSISYLDDELTTGFPSPSLYEESKS 87
DB 1 MGPSSLYSOOVLCSYVPLSKNVHGVGVCTGNIQSISYLDDELTTGFPSPSLYEESKS 60
QY 88 KEAKRELNIIVAVLNCNELLVLRKNLLAQESVETQNLKLGSDMDHLSQCYAKLKEOLET 147
DB 61 KETKRELNIIVAVLNCNELLVLRKNLLAQENVEQNLKLGSDMDHLSQCYAKLKEOLET 120
QY 148 SRREMIGLQERDRQLOCKNRSLHQLKNEKDEQVQLNTIATSRATQYNHDKKEREYNK 207
DB 121 SRREMIGLQERDRQLOCKNRSLHQLKNEKDEQVQLNTIATSRATQYNHDKKEREYNK 180
QY 208 LKERLHQLVNNKDKKIAMDILNYYGRADGKRGSWRTKTEARNEDEMYKILLNDVEYRQ 267
DB 181 LKERLHQLVNNKDKKIAMDILNYYGRADGKRGSWRTKTEARNEDEMYKILLNDVEYRQ 240
QY 268 KOILMENAEELKKVLOQMKEMLSLSPQKKPRERAEDGTGTVAISDIEDDSCGELSDSV 327
DB 241 KOILMENAEELKKVLOQMKEMLSLSPQKKPRERVDSTGTV- ISDVEDAGELRESM 299
QY 328 WGLSCDTVREQLTNSIRKQWRILKSHVEKLDNOASKVHSEGLNEEDVI SRQDHEQTEKL 387
DB 300 WDLSCETVREQLTNSIRKQWRILKSHVEKLDNOASKVHSEGLNEEDVI SRQDHEQTEKL 359
QY 388 ELEIERCKEMIRAKOQLOQLATTCDDDTTSLRDCYLLEKERLKEWTLPEKOKNF 447
DB 360 ELEIOCKEMIRAKOQLOQLATTCDDDTTSLRDCYLLEKERLKEWTLPEKOKNF 419
QY 448 ERERSFTEAAILRLGLERKAFEEERASWVKQFQFLNMTNFDHONSENKVLFSAPSGSSDPD 507
DB 420 ERERSFTEAAILRLGLERKAFEEERASWVKQFQFLNMTTDFHONSENKVLFSAPSGSSDWD 479
QY 508 NLIVHSRPRQKHLHVGVPACTSKLTSLPASPSTSDPROTHSCVSEHSSISVLNITP 567
DB 480 NLIVHSRPRQKHLHVGVPACTSKLTSLPASPSTSDPROTHSCVSEHSSISVLNITP 539
QY 568 EESKPSSEVARESTQKWSVQSPSPSREG 595
DB 540 EBKPNQVGGERTNQKWSVSPSPSREG 567
RESULT 4
ADB65313
ID ADB65313 standard; protein; 504 AA.
XX
AC ADB65313;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human protein encoded by clone TESTI20071130.
XX
KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
KW cell regeneration; membrane protein; signal transduction-related protein;
KW transcription-related protein; osteoporosis; neurological disease;
KW cancer; tumour.
XX

PA (TABA/) TABASKA J E.
XX (CAOV/) CAO Y.
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
XX Claim 1; SEQ ID NO 54149; 15pp; English.
XX
XX The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.upto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This is the amino acid sequence of a plant full length insert
CC polypeptide that can be used in the recombinant DNA construct of the
XX invention.
XX
XX Sequence 443 AA;
SQ

Query Match 12.0%; Score 379.5; DB 8; Length 443;
Best Local Similarity 26.6%; Pred. No. 3.5e-18;
Matches 106; Conservative 101; Mismatches 135; Indels 57; Gaps 14;
QY 27 KMSPSLY-----SQVLCSYPLSKNVHGVFCTGENIEQISYSLDQELTTGFPSSL 81
Db 54 RMSSARFOLRASSQOL--QPPPSHASMSSDGGAFANAENLHCARYLNQTLVTFGPPA- 110
QY 82 YVESKSKAKRLNLTAVLNCNELLVLRKNLLAQESVETONLKLGSMDHLCQCYAKL 141
Db 111 ---SLDLFATDPVSTARTCNLCYALLOQRDIEFRESTNDORQMQSDISLEAKIERM 167
QY 142 KQLETSTRREMIGLQERDRLOCKNRSLSHLLKNEKDEVQKLNQNTIASR--ATQYNHDKV 199
Db 168 DAQLAAKDELATLTRTEAKNTAALKSQIDKLQERDEFQKM--VIGNQVRTQIHEMK 225
QY 200 RKEREYNKLERLHQLVMNKKDKN--IAMDVLNYVGRADGKGSWRTDKTEARNEDEMYK 257
Db 226 KKEKEYIKLQELKNQVLMEKKKESRSRSGMEIMNLLQKEGRQGTWNGKK---NDNDYYK 281
QY 258 ILLNDYEVROKILMENAEKVKVLQMKKEMISLLSPQKKPRERAEDGTGTVAISDI-E 316
Db 282 MIVDAYEVKKQELMGENADRLALLSMQMDRDFLN-----APNGSSQSTVTDNGR 332
QY 317 DDSGEL-----SRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDN--QASKVHSEGL 369
Db 333 QESGSPQSLGGKTDVDFLPPHWARDQIBESLRTKWTISIKARMTQLQDAKQAEVTSAT 392
QY 370 NEEDVISRDHQEFTEKLELETERCKEMIKAAQQQLLQQQ 408
Db 393 DRE-----LELE-----AQLVEA-RSIIQEQ 412

RESULT 8
ID AAG40470 standard; protein; 382 AA.
XX
AC AAG40470;

XX 18-OCT-2000 (first entry)
DT
XX Arabidopsis thaliana protein fragment SEQ ID NO: 50219.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
PN
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-00301439.
PF
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 18-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.

PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	22-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	28-JUN-1999;	99US-0140991P.
PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143342P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	20-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	21-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	22-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Query Match 10.8%; Score 343; DB 3; Length 382;		
Best Local Similarity 25.8%; Pred. No. 1.1e-15;		
Matches 95; Conservative 92; Mismatches 139; Indels 42; Gaps 10;		
QY	56	FCGTGNIQSISYLDQELTTFGPPSLYEESKSKEARBLNIVAVLNCMLNLVLQKKNLL 115
Db	28	FADVDNLENCIKYLNQSLVTSGFS- ---SLDLFATDPVSIARTCNCVVALIQQRQDVE 83
QY	116	AQESVETQNLKLGSDMDHLOSCYAKLKEQLETSRREMGLQERDRLOQCNKLSHOLLKN 175
Db	84	FRESANDQORLLSDMARLEAKVERLETQLOAKERELGSVTRTEAKNTAALKTONEKLOK 143
QY	176	EKDEVOKLQNIATSR- -ATQYNHDVKREREYNKLERLHOLVM- NKDKDNIAMDVLVV 232
Db	144	ERDEFQRM- -VIANQOVTKOQLHETKKEKEVYKLOERLQVLMKKEKKEKTRSGMIMNLL 201
QY	233	GRADGKRGSRWTKTEARNEDEMYKILLNDYEYKQILMENAEKLVLOQMKKEMISLL 292
Db	202	QKEGRQRTGWSGKKTDS- ---DFYKKIVDAYEAKNQELMAENTDLRALLRSTQGDMSRSL 257
QY	293	SPQKKKPRERABDG- -----TGTVAISDIEDDSGELSRDSVWGLSCDTVREQLTNSIR 344

```
Db      258 N-----ASGLTNQSLVANGRHGADPSQSLG--GKTDVFDLPRFMARGQIEDSLR 306
Qy      345 KQWRILKSHVEKLDNQASKVHSEGLNEEDVISRQDHQETKLELEIERCKEMIKAAQOOL 404
Db      307 SKMWIKERMGLVDAQEVS-----ITSEASERELE-LEAQLVEARSIIOEQESI 356
Qy      405 LOOQLATT 412
Db      357 MSKHLPKT 364

RESULT 9
AAG48364
ID  AAG48364 standard; protein; 347 AA.
XX
AC  AAG48364;
XX
DT  18-OCT-2000 (first entry)
XX
DE  Arabidopsis thaliana protein fragment SEQ ID NO: 61066.
XX
KW  Protein identification; signal transduction pathway; metabolic pathway;
KW  hybridization assay; genetic mapping; gene expression control; promoter;
KW  termination sequence.
XX
OS  Arabidopsis thaliana.
XX
PN  EP1033405-A2.
XX
PD
PP  06-SEP-2000.
XX
XX  25-FEB-2000; 2000EP-00301439.
XX
XX  25-FEB-1999; 99US-0121825P.
PR  05-MAR-1999; 99US-0123180P.
PR  09-MAR-1999; 99US-0123548P.
PR  23-MAR-1999; 99US-0125788P.
PR  25-MAR-1999; 99US-0126264P.
PR  29-MAR-1999; 99US-0126785P.
PR  01-APR-1999; 99US-0127462P.
PR  06-APR-1999; 99US-0128234P.
PR  08-APR-1999; 99US-0128714P.
PR  16-APR-1999; 99US-0129845P.
PR  19-APR-1999; 99US-0130077P.
PR  21-APR-1999; 99US-0130449P.
PR  23-APR-1999; 99US-0130510P.
PR  28-APR-1999; 99US-0130891P.
PR  30-APR-1999; 99US-0131449P.
PR  30-APR-1999; 99US-0132048P.
PR  04-MAY-1999; 99US-0132407P.
PR  05-MAY-1999; 99US-0132484P.
PR  06-MAY-1999; 99US-0132485P.
PR  06-MAY-1999; 99US-0132486P.
PR  07-MAY-1999; 99US-0132487P.
PR  11-MAY-1999; 99US-013256P.
PR  14-MAY-1999; 99US-0134218P.
PR  14-MAY-1999; 99US-0134219P.
PR  14-MAY-1999; 99US-0134221P.
PR  14-MAY-1999; 99US-0134370P.
PR  18-MAY-1999; 99US-0134768P.
PR  19-MAY-1999; 99US-0134941P.
PR  20-MAY-1999; 99US-0135124P.
PR  21-MAY-1999; 99US-0135353P.
PR  24-MAY-1999; 99US-0135629P.
PR  25-MAY-1999; 99US-0136021P.
PR  27-MAY-1999; 99US-0136392P.
PR  28-MAY-1999; 99US-0136782P.
PR  01-JUN-1999; 99US-0137222P.
PR  03-JUN-1999; 99US-0137528P.
PR  04-JUN-1999; 99US-0137502P.
PR  07-JUN-1999; 99US-0137724P.
PR  08-JUN-1999; 99US-0138094P.
PR  10-JUN-1999; 99US-0138540P.
PR  10-JUN-1999; 99US-0138847P.
PR  14-JUN-1999; 99US-0139119P.
PR  16-JUN-1999; 99US-0139452P.
PR  16-JUN-1999; 99US-0139453P.
PR  17-JUN-1999; 99US-0139492P.
PR  18-JUN-1999; 99US-0139454P.
PR  18-JUN-1999; 99US-0139455P.
PR  18-JUN-1999; 99US-0139456P.
PR  18-JUN-1999; 99US-0139457P.
PR  18-JUN-1999; 99US-0139458P.
PR  18-JUN-1999; 99US-0139459P.
PR  18-JUN-1999; 99US-0139460P.
PR  18-JUN-1999; 99US-0139461P.
PR  18-JUN-1999; 99US-0139462P.
PR  18-JUN-1999; 99US-0139463P.
PR  18-JUN-1999; 99US-0139750P.
PR  18-JUN-1999; 99US-0139763P.
PR  21-JUN-1999; 99US-0139817P.
PR  22-JUN-1999; 99US-0139899P.
PR  23-JUN-1999; 99US-0140353P.
PR  23-JUN-1999; 99US-0140354P.
PR  24-JUN-1999; 99US-0140695P.
PR  28-JUN-1999; 99US-0140823P.
PR  29-JUN-1999; 99US-0140991P.
PR  30-JUN-1999; 99US-0141287P.
PR  01-JUL-1999; 99US-0141842P.
PR  01-JUL-1999; 99US-0142154P.
PR  02-JUL-1999; 99US-0142055P.
PR  06-JUL-1999; 99US-0142390P.
PR  08-JUL-1999; 99US-0142803P.
PR  09-JUL-1999; 99US-0142920P.
PR  12-JUL-1999; 99US-0142977P.
PR  13-JUL-1999; 99US-0143542P.
PR  14-JUL-1999; 99US-0143624P.
PR  15-JUL-1999; 99US-0144005P.
PR  16-JUL-1999; 99US-0144085P.
PR  16-JUL-1999; 99US-0144086P.
PR  19-JUL-1999; 99US-0144325P.
PR  19-JUL-1999; 99US-0144331P.
PR  19-JUL-1999; 99US-0144332P.
PR  19-JUL-1999; 99US-0144333P.
PR  19-JUL-1999; 99US-0144334P.
PR  19-JUL-1999; 99US-0144335P.
PR  20-JUL-1999; 99US-0144352P.
PR  20-JUL-1999; 99US-0144632P.
PR  20-JUL-1999; 99US-0144884P.
PR  20-JUL-1999; 99US-0144814P.
PR  21-JUL-1999; 99US-0145086P.
PR  21-JUL-1999; 99US-0145088P.
PR  22-JUL-1999; 99US-0145085P.
PR  22-JUL-1999; 99US-0145087P.
PR  22-JUL-1999; 99US-0145089P.
PR  22-JUL-1999; 99US-0145192P.
PR  23-JUL-1999; 99US-0145145P.
PR  23-JUL-1999; 99US-0145218P.
PR  23-JUL-1999; 99US-0145224P.
PR  26-JUL-1999; 99US-0145276P.
PR  27-JUL-1999; 99US-0145913P.
PR  27-JUL-1999; 99US-0145918P.
PR  27-JUL-1999; 99US-0145919P.
PR  28-JUL-1999; 99US-0145951P.
PR  02-AUG-1999; 99US-0146386P.
PR  02-AUG-1999; 99US-0146388P.
PR  02-AUG-1999; 99US-0146389P.
PR  03-AUG-1999; 99US-0147038P.
PR  04-AUG-1999; 99US-0147204P.
PR  04-AUG-1999; 99US-0147302P.
PR  05-AUG-1999; 99US-0147192P.
PR  05-AUG-1999; 99US-0147260P.
PR  06-AUG-1999; 99US-0147303P.
PR  06-AUG-1999; 99US-0147416P.
```

PR	09-AUG-1999;	99US-0147493P.	QY	56	FCTGENIEQISISYLDQELTTFGFPSPSYBESKSKAKRELNIVAVLNCMELLVLQKNLL	115
PR	09-AUG-1999;	99US-0147935P.				
PR	10-AUG-1999;	99US-0148171P.	DB	24	FANEDNLEHCTKYLNTQMTVTFGPPA-----SLDLFSDNPVSIISRTCNMYSLQQRORDIE	79
PR	11-AUG-1999;	99US-0148311P.				
PR	12-AUG-1999;	99US-0148341P.	QY	116	AOESVETQNLKLGSDMDHLSQCVAKLKEOLETSSREMIQLQSRDRLOLCKNSLHOLLKN	175
PR	13-AUG-1999;	99US-0148565P.				
PR	13-AUG-1999;	99US-0148684P.	DB	80	FRESANELRQROQSDIARLEAKVERLEALLQOKDRDEIATITTEAKNTAALKSQIEKLOQ	139
PR	16-AUG-1999;	99US-0149368P.				
PR	17-AUG-1999;	99US-0149175P.				
PR	18-AUG-1999;	99US-0149426P.	QY	176	EKDEVOKLQNIISAR--ATQYNHDVKRKEREYNKLERLHQLVM-NKKDKNIAMDVLNVY	232
PR	20-AUG-1999;	99US-0149722P.	DB	140	ERDEFORM--VIGNQOVKAQIHEMKKEKDYIKQLERLNVLMKKEKESRSRSMETMULL	197
PR	20-AUG-1999;	99US-0149723P.				
PR	20-AUG-1999;	99US-0149929P.	QY	233	GRADGKRGSRWTDKTEARNEDEMYKILLNDIYRQKQIILMENAEKLVVLOQMKEMISLL	292
PR	23-AUG-1999;	99US-0149902P.				
PR	23-AUG-1999;	99US-0149930P.	DB	198	OKEGRQGRGTWNGKXTDT-----DFYKKIVDAYEAKNQELMAENTSRLALLRSQOTDMRDL	253
PR	25-AUG-1999;	99US-0150566P.				
PR	26-AUG-1999;	99US-0150884P.	QY	293	SPQKKPRERAEDGTGTVAISDIEDDSGELSDSDSVGLSGCDTVREQLTNSIRKOWILKS	352
PR	27-AUG-1999;	99US-0151065P.				
PR	27-AUG-1999;	99US-0151066P.	DB	254	N-----APNGSATLAGS--EKREADPSOSPGGKTESMV--QLQDAPKR-----	293
PR	27-AUG-1999;	99US-0151080P.				
PR	30-AUG-1999;	99US-0151303P.				
PR	31-AUG-1999;	99US-0151438P.	QY	353	HVEKLDNQASKVHSEGLNEEDVISQDHQETKELEIERCKEMIKAOQQLLQOOL	409
PR	01-SEP-1999;	99US-0151930P.	DB	294	-----ASVTSEAT-----ERELE-LEAQLVEARSIIQEQESIMSKHL	329
PR	07-SEP-1999;	99US-0152363P.				
PR	10-SEP-1999;	99US-0153070P.				
PR	13-SEP-1999;	99US-0153758P.				
PR	16-SEP-1999;	99US-0154018P.	RESULT 10			
PR	16-SEP-1999;	99US-0154039P.	ADX76571			
PR	20-SEP-1999;	99US-0154779P.	XX	ID	ADX76571 standard; protein; 290 AA.	
PR	23-SEP-1999;	99US-0155139P.	AC	ADX76571;		
PR	23-SEP-1999;	99US-0155486P.	XX			
PR	24-SEP-1999;	99US-0155659P.	DT	23-MAR-2006 (revised)		
PR	28-SEP-1999;	99US-0156458P.	DT	21-APR-2005 (first entry)		
PR	29-SEP-1999;	99US-0156596P.	XX			
PR	04-OCT-1999;	99US-0157117P.	DE	Plant full length insert polypeptide seqid 45937.		
PR	05-OCT-1999;	99US-0157753P.	XX			
PR	06-OCT-1999;	99US-0157865P.	XX	plant protectant; plant growth regulant; gene therapy; plant;		
PR	07-OCT-1999;	99US-0158029P.	KW	recombinant DNA construct; physical array; plant breeding marker;		
PR	08-OCT-1999;	99US-0158232P.	KW	cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;		
PR	12-OCT-1999;	99US-0158369P.	KW	extreme osmotic condition; pathogen tolerance; pest tolerance;		
PR	13-OCT-1999;	99US-0159293P.	KW	growth rate; cell cycle pathway; disease resistance;		
PR	13-OCT-1999;	99US-0159294P.	KW	galactomannan production; lignin production; plant growth regulator;		
PR	13-OCT-1999;	99US-0159295P.	KW	yield; plant growth; plant development; seed oil; protein yield;		
PR	14-OCT-1999;	99US-0159329P.	KW	protein content.		
PR	14-OCT-1999;	99US-0159330P.	XX			
PR	14-OCT-1999;	99US-0159331P.	OS	Glycine max.		
PR	14-OCT-1999;	99US-0159637P.	XX			
PR	14-OCT-1999;	99US-0159638P.	PN	US2004034888-A1.		
PR	18-OCT-1999;	99US-0159584P.	XX			
PR	21-OCT-1999;	99US-0160741P.	PD	19-FEB-2004.		
PR	21-OCT-1999;	99US-0160767P.	XX			
PR	21-OCT-1999;	99US-0160768P.	PF	28-APR-2003; 2003US-00425114.		
PR	21-OCT-1999;	99US-0160770P.	XX			
PR	21-OCT-1999;	99US-0160814P.	PR	06-MAY-1999; 99US-00304517.		
PR	21-OCT-1999;	99US-0160815P.	PR	05-NOV-2001; 2001US-00985678.		
PR	22-OCT-1999;	99US-0160980P.	XX			
PR	22-OCT-1999;	99US-0160981P.	PA	(LIUJ/) LIU J.		
PR	22-OCT-1999;	99US-0160989P.	PA	(ZHOU/) ZHOU Y.		
PR	25-OCT-1999;	99US-0161404P.	PA	(KOVA/) KOVALIC D K.		
PR	25-OCT-1999;	99US-0161405P.	PA	(SCRE/) SCREEN S E.		
PR	25-OCT-1999;	99US-0161406P.	PA	(TABAJ/) TABASKA J E.		
PR	26-OCT-1999;	99US-0161359P.	PA	(CAOY/) CAO Y.		
PR	26-OCT-1999;	99US-0161360P.	XX			
PR	26-OCT-1999;	99US-0161361P.	PI	Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;		
PR	28-OCT-1999;	99US-0161920P.	XX			
PR	28-OCT-1999;	99US-0161992P.	DR	WPI; 2004-180133/17.		
PR	28-OCT-1999;	99US-0161993P.	XX			
PR	29-OCT-1999;	99US-0162142P.	PT	New recombinant DNA construct, useful for improving plant tolerance to		
			PT	cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or		
			PT	pests, for conferring increased resistance to plant disease, or for		
			PT	improving yield.		
			PS	Claim 1; SEQ ID NO 45937; 15pp; English.		

Query Match 9.8%; Score 310; DB 3; Length 347;
Best Local Similarity 24.9%; Pred. No. 2.le-13;
Matches 89; Conservative 89; Mismatches 125; Indels 54; Gaps 11;

PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	05-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	06-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	09-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	16-SEP-1999;	99US-0154018P.
PR	20-SEP-1999;	99US-0154039P.
PR	22-SEP-1999;	99US-0154779P.
PR	23-SEP-1999;	99US-0155139P.
PR	24-SEP-1999;	99US-0155486P.
PR	28-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	28-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Query Match 8.8%; Score 277; DB 3; Length 291;		
Best Local Similarity 25.7%; Pred. No. 3 6e-11;		
Matches 75; Conservative 81; Mismatches 90; Indels 46; Gaps 11;		
QY	129	SDMHLCYAKLKEQLETSRREMIGLQERDRQLQCKNRSLHOLLKNEKDEVKLQNIITA 188
Db	3	SDISLEAKIERMDAQLAAKDRELATLTRXEAKNATAALKSQIDKLQQRDERFQKM--VIG 60
QY	189	SR--ATOYNHDVKREREYKLERLHQLVNNKKDKN--IANDVLNVGRADGKGSWRT 244
Db	61	NOQVETQOIHEMKKEKEYIKLOEKLNOVLMEKKESSESGMEIWNLLQKGRQRTWNG 120
QY	245	DKTEARNEDEMYKILLINDYEVROKQILMENAEKLVLOQMKEMISLLSPOKKKPRERAE 304
Db	121	KK----NDNDYYKMIVDAYEVKKQELMQENADLRALLRSQMQMDRDLN-----AP 167
QY	305	DGTGTVASDI-EDDSGEL-----SRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLD 358
Db	168	NGSSQSTVTDNGRQESGPSQSPLGKGTVDVDFLPPHWARDQIESLRTKMTSIKARMTQLQ 227
QY	359	N--QASKVHSEGLNEEDVISRODHEQETEKLEIERCKEMIKAAQOQLLOQQ 408
Db	228	DAQKGAEVTSEATDRE-----LELE----AOLVEA-RSIIQEQ 260
RESULT 12		
AAG48365		
ID	AAG48365	standard; protein; 307 AA.
XX	AC	AAG48365;
XX	18-OCT-2000	(first entry)
DT	18-OCT-2000	(first entry)
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 61067.
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX	Arabidopsis thaliana.	
OS	Arabidopsis thaliana.	

XX EP1033405-A2.
PN 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0128645P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0132566P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147453P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149436P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151343P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.

PR 18-JUN-1999;	99US-0139461P.	PR 23-AUG-1999;	99US-0149930P.
PR 18-JUN-1999;	99US-0139462P.	PR 25-AUG-1999;	99US-0150566P.
PR 18-JUN-1999;	99US-0139463P.	PR 26-AUG-1999;	99US-0150884P.
PR 18-JUN-1999;	99US-0139750P.	PR 27-AUG-1999;	99US-0151065P.
PR 18-JUN-1999;	99US-0139763P.	PR 27-AUG-1999;	99US-0151066P.
PR 21-JUN-1999;	99US-0139817P.	PR 27-AUG-1999;	99US-0151080P.
PR 22-JUN-1999;	99US-0139899P.	PR 30-AUG-1999;	99US-0151303P.
PR 23-JUN-1999;	99US-0140353P.	PR 31-AUG-1999;	99US-0151438P.
PR 23-JUN-1999;	99US-0140354P.	PR 01-SEP-1999;	99US-0151930P.
PR 24-JUN-1999;	99US-0140695P.	PR 07-SEP-1999;	99US-0152363P.
PR 28-JUN-1999;	99US-0140823P.	PR 10-SEP-1999;	99US-0153070P.
PR 29-JUN-1999;	99US-0140991P.	PR 13-SEP-1999;	99US-0153758P.
PR 30-JUN-1999;	99US-0141287P.	PR 15-SEP-1999;	99US-0154018P.
PR 01-JUL-1999;	99US-0141842P.	PR 16-SEP-1999;	99US-0154039P.
PR 01-JUL-1999;	99US-0142154P.	PR 20-SEP-1999;	99US-0154779P.
PR 02-JUL-1999;	99US-0142055P.	PR 22-SEP-1999;	99US-0155139P.
PR 06-JUL-1999;	99US-0142390P.	PR 23-SEP-1999;	99US-0155486P.
PR 08-JUL-1999;	99US-0142803P.	PR 24-SEP-1999;	99US-0155659P.
PR 09-JUL-1999;	99US-0142920P.	PR 28-SEP-1999;	99US-0156458P.
PR 12-JUL-1999;	99US-0142977P.	PR 29-SEP-1999;	99US-0156596P.
PR 13-JUL-1999;	99US-0143542P.	PR 04-OCT-1999;	99US-0157117P.
PR 14-JUL-1999;	99US-0143624P.	PR 05-OCT-1999;	99US-0157753P.
PR 15-JUL-1999;	99US-0144005P.	PR 06-OCT-1999;	99US-0157865P.
PR 16-JUL-1999;	99US-0144085P.	PR 07-OCT-1999;	99US-0158029P.
PR 16-JUL-1999;	99US-0144086P.	PR 08-OCT-1999;	99US-0158232P.
PR 19-JUL-1999;	99US-0144325P.	PR 12-OCT-1999;	99US-0158369P.
PR 19-JUL-1999;	99US-0144331P.	PR 13-OCT-1999;	99US-0159293P.
PR 19-JUL-1999;	99US-0144332P.	PR 13-OCT-1999;	99US-0159294P.
PR 19-JUL-1999;	99US-0144333P.	PR 13-OCT-1999;	99US-0159295P.
PR 19-JUL-1999;	99US-0144334P.	PR 14-OCT-1999;	99US-0159329P.
PR 19-JUL-1999;	99US-0144335P.	PR 14-OCT-1999;	99US-0159330P.
PR 20-JUL-1999;	99US-0144352P.	PR 14-OCT-1999;	99US-0159331P.
PR 20-JUL-1999;	99US-0144632P.	PR 14-OCT-1999;	99US-0159637P.
PR 20-JUL-1999;	99US-0144884P.	PR 14-OCT-1999;	99US-0159638P.
PR 21-JUL-1999;	99US-0144814P.	PR 18-OCT-1999;	99US-0159584P.
PR 21-JUL-1999;	99US-0145086P.	PR 21-OCT-1999;	99US-0160741P.
PR 21-JUL-1999;	99US-0145088P.	PR 21-OCT-1999;	99US-0160767P.
PR 22-JUL-1999;	99US-0145085P.	PR 21-OCT-1999;	99US-0160768P.
PR 22-JUL-1999;	99US-0145087P.	PR 21-OCT-1999;	99US-0160770P.
PR 22-JUL-1999;	99US-0145089P.	PR 21-OCT-1999;	99US-0160814P.
PR 22-JUL-1999;	99US-0145192P.	PR 21-OCT-1999;	99US-0160815P.
PR 23-JUL-1999;	99US-0145145P.	PR 22-OCT-1999;	99US-0160980P.
PR 23-JUL-1999;	99US-0145218P.	PR 22-OCT-1999;	99US-0160981P.
PR 23-JUL-1999;	99US-0145224P.	PR 22-OCT-1999;	99US-0160989P.
PR 26-JUL-1999;	99US-0145276P.	PR 25-OCT-1999;	99US-0161404P.
PR 27-JUL-1999;	99US-0145913P.	PR 25-OCT-1999;	99US-0161405P.
PR 27-JUL-1999;	99US-0145918P.	PR 25-OCT-1999;	99US-0161406P.
PR 27-JUL-1999;	99US-0145919P.	PR 26-OCT-1999;	99US-0161359P.
PR 28-JUL-1999;	99US-0145951P.	PR 26-OCT-1999;	99US-0161360P.
PR 02-AUG-1999;	99US-0146386P.	PR 26-OCT-1999;	99US-0161361P.
PR 02-AUG-1999;	99US-0146388P.	PR 28-OCT-1999;	99US-0161920P.
PR 02-AUG-1999;	99US-0146389P.	PR 28-OCT-1999;	99US-0161992P.
PR 03-AUG-1999;	99US-0147038P.	PR 28-OCT-1999;	99US-0161993P.
PR 04-AUG-1999;	99US-0147204P.	PR 29-OCT-1999;	99US-0162142P.
PR 04-AUG-1999;	99US-0147302P.		
PR 05-AUG-1999;	99US-0147192P.		
PR 05-AUG-1999;	99US-0147260P.		
PR 06-AUG-1999;	99US-0147303P.		
PR 06-AUG-1999;	99US-0147416P.		
PR 09-AUG-1999;	99US-0147933P.		
PR 09-AUG-1999;	99US-0147935P.		
PR 10-AUG-1999;	99US-0148171P.		
PR 11-AUG-1999;	99US-0148319P.		
PR 12-AUG-1999;	99US-0148341P.		
PR 13-AUG-1999;	99US-0148565P.		
PR 13-AUG-1999;	99US-0148684P.		
PR 16-AUG-1999;	99US-0149368P.		
PR 17-AUG-1999;	99US-0149175P.		
PR 18-AUG-1999;	99US-0149426P.		
PR 20-AUG-1999;	99US-0149722P.		
PR 20-AUG-1999;	99US-0149723P.		
PR 23-AUG-1999;	99US-0149902P.		
<div>Query Match 8.4%; Score 265.5; DB 3; Length 288; Best Local Similarity 25.7%; Pred. No. 2.3e-10; Matches 75; Conservative 80; Mismatches 92; Indels 45; Gaps 12;</div>			
QY 135	QSCYAKLKEQLETSRREMIGLOERDQL-----QCKN-----RSLHQLLNKNEKDEVQKLQ 184		
Db 7	QSDIARLEAKVE--RLEAL-LQKDRBTATITRTEAKNTAALKSQIEKLQQRDEFORM- 62		
QY 185	NIASR--ATQYNHDVKKEREYNKIKERLHOLVM-NKKDKNIANDVLNYYVRADGKRG 241		
Db 63	-VIGNQVKAQOIHEMKKKKXDIYKLOERLQVLMKKKESRSGMEIWNLLQKEGRQRT 121		
QY 242	WRTDKTEARNEDEMYKILLNDYERQKQILMENAEKLVQOMKKEMISLLSPQKKPRE 301		
Db 122	WNGKKTDT---DFYKIVDAYEAKNQELMAENTSRLALLRSMQTDMDRDLN----- 169		
QY 302	RAEDGTGTVAJSDIEDDSGELS-----RDSVNGLSCTVTRQELTNSIRKQWRLLKSHVKL 357		

Db 170 -APNGSATLAGEKHEADPSQPLGGKTDVFDLPYRMARQIEESLRTKMSIKESMVQL 228
QY 358 DNOASKVHSEGLNEEDVIRQDHEQTEKLEIEIERCKEMIKAQOOLLOOOL 409
Db 229 QDAPKRA-----VTSEATERELE-LEAQLVEARSIIORQESIMSKHL 270

RESULT 14
AAG33166
ID AAG33166 standard; protein; 277 AA.
XX AC AAG33166;
XX DT 18-OCT-2000 (first entry)
XX DE Zea mays protein fragment SEQ ID NO: 40145.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX OS Zea mays subsp. mays.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.

PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0158458P.
PR	29-SEP-1999;	99US-0158596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160811P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.

Query Match

Best Local Similarity 8.3%; Score 262.5; DB 3; Length 277;

Mismatches 73; Conservative 73; Mismatches 70; Indels 63; Gaps 12;

QY	159	DFLOCKRSLHQL-----LKNKEKDEVOKLONLIASR--ATQYNHDVK 199
DB	2	DAQLAAKORLATLFRXEAKNTAALKSQIDKLQQRDEBFQM--VIGNQVRTQOIHEMK 59
QY	200	RKEREYNKLKERLHQLVMNKDKN--IAMDVLNLYVGRADGKRGSWRTDKTEARNEDEMYK 257

Db	60	KKEKEYIKLQKLNQVLMEKKESRSRSGMEIMNLLQKEGRQGTWNGKK---NDNDYYK 115
QY	258	ILLNDYERYRQKILMENAEKLVLOOMKKEMISLLSPQKKPRERAEDGTGTVAISDI-E 316
Db	116	MIVDAYEYKQELMQENADLRALLRSMQMDRDFLN-----APNGSSQSTVTDNGR 166
QY	317	DDSGEL-----SRDSVMGLSCDVTREQLTNSIRKQWRILKSHVEKLDN--QASKVHSEGL 369
Db	167	QESGSPQSLGGKTDVFDLPFHMAHQIEESLRTKMTSIKARMTQLQDAQKGAETVSEAT 226
QY	370	NEEDVISRODHQETEKLEIEIERCKEMIKAAQQQLLQQQ 408
Db	227	DRE-----LELE-----AQLVEA-RSIIQEQ 246
RESULT 15		
AAG40471		
ID	AAG40471	standard; protein; 284 AA.
XX	XX	AAG40471;
XX	XX	18-OCT-2000 (first entry)
XX	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 50220.
XX	XX	Protein identification; signal transduction pathway; metabolic pathway;
XX	XX	hybridisation assay; genetic mapping; gene expression control; promoter;
XX	XX	termination sequence.
XX	XX	Arabidopsis thaliana.
XX	XX	EP1033405-A2.
PD	06-SEP-2000.	
XX	25-FEB-2000;	2000EP-00301439.
PR	25-FEB-1999;	99US-0121825P.
PR	05-MAR-1999;	99US-0123180P.
PR	09-MAR-1999;	99US-0123548P.
PR	23-MAR-1999;	99US-0125788P.
PR	25-MAR-1999;	99US-0126264P.
PR	29-MAR-1999;	99US-0126785P.
PR	01-APR-1999;	99US-0127462P.
PR	06-APR-1999;	99US-0128234P.
PR	08-APR-1999;	99US-0128714P.
PR	16-APR-1999;	99US-0129845P.
PR	19-APR-1999;	99US-0130077P.
PR	21-APR-1999;	99US-0130449P.
PR	23-APR-1999;	99US-0130510P.
PR	28-APR-1999;	99US-0130891P.
PR	30-APR-1999;	99US-0131449P.
PR	30-APR-1999;	99US-0132048P.
PR	04-MAY-1999;	99US-0132407P.
PR	05-MAY-1999;	99US-0132484P.
PR	05-MAY-1999;	99US-0132485P.
PR	06-MAY-1999;	99US-0132486P.
PR	06-MAY-1999;	99US-0132487P.
PR	07-MAY-1999;	99US-0132863P.
PR	11-MAY-1999;	99US-0134256P.
PR	14-MAY-1999;	99US-0134218P.
PR	14-MAY-1999;	99US-0134219P.
PR	14-MAY-1999;	99US-0134221P.
PR	14-MAY-1999;	99US-0134370P.
PR	18-MAY-1999;	99US-0134768P.
PR	19-MAY-1999;	99US-0134941P.
PR	20-MAY-1999;	99US-0135124P.
PR	21-MAY-1999;	99US-0135353P.
PR	24-MAY-1999;	99US-0135629P.
PR	25-MAY-1999;	99US-0136021P.
PR	27-MAY-1999;	99US-0136392P.
PR	28-MAY-1999;	99US-0136782P.

PR 01-JUN-1999;	99US-01373222P.	PR 05-AUG-1999;	99US-01471192P.
PR 03-JUN-1999;	99US-01375288P.	PR 05-AUG-1999;	99US-0147260P.
PR 04-JUN-1999;	99US-01375028P.	PR 06-AUG-1999;	99US-0147303P.
PR 07-JUN-1999;	99US-0137724P.	PR 06-AUG-1999;	99US-0147416P.
PR 08-JUN-1999;	99US-0138094P.	PR 09-AUG-1999;	99US-0147493P.
PR 10-JUN-1999;	99US-0138540P.	PR 09-AUG-1999;	99US-0147935P.
PR 10-JUN-1999;	99US-0138847P.	PR 10-AUG-1999;	99US-0148171P.
PR 14-JUN-1999;	99US-0139119P.	PR 11-AUG-1999;	99US-0148319P.
PR 16-JUN-1999;	99US-0139452P.	PR 12-AUG-1999;	99US-0148341P.
PR 17-JUN-1999;	99US-0139453P.	PR 13-AUG-1999;	99US-0148565P.
PR 18-JUN-1999;	99US-0139492P.	PR 13-AUG-1999;	99US-0148565P.
PR 18-JUN-1999;	99US-0139454P.	PR 16-AUG-1999;	99US-0148684P.
PR 18-JUN-1999;	99US-0139455P.	PR 17-AUG-1999;	99US-0149175P.
PR 18-JUN-1999;	99US-0139456P.	PR 18-AUG-1999;	99US-0149266P.
PR 18-JUN-1999;	99US-0139457P.	PR 20-AUG-1999;	99US-0149722P.
PR 18-JUN-1999;	99US-0139458P.	PR 20-AUG-1999;	99US-0149723P.
PR 18-JUN-1999;	99US-0139459P.	PR 20-AUG-1999;	99US-0149929P.
PR 18-JUN-1999;	99US-0139460P.	PR 23-AUG-1999;	99US-0149902P.
PR 18-JUN-1999;	99US-0139461P.	PR 23-AUG-1999;	99US-0149930P.
PR 18-JUN-1999;	99US-0139462P.	PR 25-AUG-1999;	99US-0150566P.
PR 18-JUN-1999;	99US-0139463P.	PR 26-AUG-1999;	99US-0150884P.
PR 18-JUN-1999;	99US-0139750P.	PR 27-AUG-1999;	99US-0151065P.
PR 18-JUN-1999;	99US-0139763P.	PR 27-AUG-1999;	99US-0151066P.
PR 21-JUN-1999;	99US-0139817P.	PR 27-AUG-1999;	99US-0151080P.
PR 22-JUN-1999;	99US-0139899P.	PR 30-AUG-1999;	99US-0151303P.
PR 23-JUN-1999;	99US-0140353P.	PR 31-AUG-1999;	99US-0151438P.
PR 23-JUN-1999;	99US-0140354P.	PR 01-SEP-1999;	99US-0151930P.
PR 24-JUN-1999;	99US-0140695P.	PR 07-SEP-1999;	99US-0152363P.
PR 28-JUN-1999;	99US-0140823P.	PR 10-SEP-1999;	99US-0153070P.
PR 29-JUN-1999;	99US-0140991P.	PR 13-SEP-1999;	99US-0153758P.
PR 30-JUN-1999;	99US-0141287P.	PR 15-SEP-1999;	99US-0154018P.
PR 01-JUL-1999;	99US-0141842P.	PR 16-SEP-1999;	99US-0154039P.
PR 01-JUL-1999;	99US-0142154P.	PR 20-SEP-1999;	99US-0155139P.
PR 02-JUL-1999;	99US-0142055P.	PR 22-SEP-1999;	99US-0154779P.
PR 06-JUL-1999;	99US-0142390P.	PR 23-SEP-1999;	99US-0155486P.
PR 08-JUL-1999;	99US-0142803P.	PR 24-SEP-1999;	99US-0155659P.
PR 09-JUL-1999;	99US-0142920P.	PR 28-SEP-1999;	99US-0156458P.
PR 12-JUL-1999;	99US-0142977P.	PR 29-SEP-1999;	99US-0156596P.
PR 13-JUL-1999;	99US-0143542P.	PR 04-OCT-1999;	99US-0157117P.
PR 14-JUL-1999;	99US-0143624P.	PR 05-OCT-1999;	99US-0157753P.
PR 15-JUL-1999;	99US-0144005P.	PR 06-OCT-1999;	99US-0157865P.
PR 16-JUL-1999;	99US-0144085P.	PR 07-OCT-1999;	99US-0158029P.
PR 16-JUL-1999;	99US-0144086P.	PR 08-OCT-1999;	99US-0158232P.
PR 19-JUL-1999;	99US-0144325P.	PR 12-OCT-1999;	99US-0158369P.
PR 19-JUL-1999;	99US-0144331P.	PR 13-OCT-1999;	99US-0159293P.
PR 19-JUL-1999;	99US-0144332P.	PR 13-OCT-1999;	99US-0159294P.
PR 19-JUL-1999;	99US-0144333P.	PR 13-OCT-1999;	99US-0159295P.
PR 19-JUL-1999;	99US-0144334P.	PR 14-OCT-1999;	99US-0159329P.
PR 20-JUL-1999;	99US-0144335P.	PR 14-OCT-1999;	99US-0159330P.
PR 20-JUL-1999;	99US-0144352P.	PR 14-OCT-1999;	99US-0159331P.
PR 20-JUL-1999;	99US-0144632P.	PR 14-OCT-1999;	99US-0159637P.
PR 21-JUL-1999;	99US-0144884P.	PR 14-OCT-1999;	99US-0159638P.
PR 21-JUL-1999;	99US-0144814P.	PR 18-OCT-1999;	99US-0159584P.
PR 21-JUL-1999;	99US-0145086P.	PR 21-OCT-1999;	99US-0160741P.
PR 21-JUL-1999;	99US-0145088P.	PR 21-OCT-1999;	99US-0160767P.
PR 22-JUL-1999;	99US-0145085P.	PR 21-OCT-1999;	99US-0160768P.
PR 22-JUL-1999;	99US-0145087P.	PR 21-OCT-1999;	99US-0160770P.
PR 22-JUL-1999;	99US-0145089P.	PR 21-OCT-1999;	99US-0160814P.
PR 23-JUL-1999;	99US-0145192P.	PR 21-OCT-1999;	99US-0160815P.
PR 23-JUL-1999;	99US-0145145P.	PR 22-OCT-1999;	99US-0160800P.
PR 23-JUL-1999;	99US-0145218P.	PR 22-OCT-1999;	99US-0160981P.
PR 23-JUL-1999;	99US-0145224P.	PR 22-OCT-1999;	99US-0160989P.
PR 26-JUL-1999;	99US-0145276P.	PR 25-OCT-1999;	99US-0161404P.
PR 27-JUL-1999;	99US-0145913P.	PR 25-OCT-1999;	99US-0161405P.
PR 27-JUL-1999;	99US-0145918P.	PR 25-OCT-1999;	99US-0161406P.
PR 27-JUL-1999;	99US-0145919P.	PR 26-OCT-1999;	99US-0161359P.
PR 28-JUL-1999;	99US-0145951P.	PR 26-OCT-1999;	99US-0161360P.
PR 02-AUG-1999;	99US-0146386P.	PR 26-OCT-1999;	99US-0161361P.
PR 02-AUG-1999;	99US-0146388P.	PR 28-OCT-1999;	99US-0161920P.
PR 02-AUG-1999;	99US-0146389P.	PR 28-OCT-1999;	99US-0161992P.
PR 03-AUG-1999;	99US-0147038P.	PR 28-OCT-1999;	99US-0161993P.
PR 04-AUG-1999;	99US-0147204P.	PR 29-OCT-1999;	99US-0162142P.
PR 04-AUG-1999;	99US-0147302P.		


```
source 1. 2692
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
80. - 1927
CDS /codon_start=1
/product="afadin- and alpha-actinin-binding protein ADIP"
/protein_id="AAO15015.1"
/db_xref="GI:27451839"
/translation="MGDMVTVDPLCTENKNLSQYTSRKMSPSSLYSQVLCSSVP
LSKNVGVGVCFGENIEQISLYIDDELTFPGPSPYEESSKEAKEALNLIIVAVLNC
MNEVLVLRNLLAQSVEVTQNLKLGSDMDHLQSCYAKLKEOLETSREMIGLQERDR
QLQCKNRLSHOLLKNEDEQVQLNIIASRATQYNHVDVKREYKYLKERLHQLVMN
LKDNIAWDVINYVGRADGKRSWRDTEARNEDENYKILLNDIETFRKQIGLMENAE
LKKVLOQMKEVMSLLSPQKKPREAEDGTGTVAISDIEDDSGELSRSVWGLSCDT
VREQLTMSIKQWILKSHVEKLDNQASKHSEGLNEEDVISRDHQEETEKLELEIE
RCKMTIAQOOLLQOOLATCDDDTTSLLRDCYLLBEKERLKEWTTFKEQKKNFIRE
RRSFTEAIIRLGLERKAFERERASWVKQOFLNTPDNPHONSENVKLFSAFGSSDPDN
LIVHSRPRQKLLHSAVANGVPACTSKLTKSLPASPSSTDSFRQTHSCVSEHSISVINIT
PEESKPSEVARESTDQRKWSVQSRPSREGCTSGCCSAFRSAHGRDDLP"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 2692
Score: 3165.00 Matches: 615
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0
US-10-644-084-2 (1-615) x AF532969 (1-2692)
Qy 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAenLysAenLeuSer 20
Db 80 ATGGAGATTGGATGACTGTGACAGATCCAGTCTGTGTACAGAAAACAAAACTCTCT 139
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLysThrGlnGlnValLeuCys 40
Db 140 CAATATACCTCAGAAACAAAGATGTCTCCGTCCTAGTTGTACTCCACGAAAGTTCTGTGC 199
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 200 TCTTCAGTACCTTTATCCAAAAACGTCATGCTGTCTTTCGGTGTCTTCGACAGAGAG 259
Qy 61 AsnIleGluGlnSerIleSerTyLysAspGlnGluLeuThrThrPheGlyPheProSer 80
Db 260 AACATTGAACAAAGTATTCTCTATCTTGTATCAGGAGCTGACCACCTTCGGGTTTCTTCC 319
Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAenIleValAlaValLeu 100
Db 320 TTGTATGAAGAAATCCAAAAGTAAAGAGCAAGAGAGAAATTAATATAGTCGCTGTCTG 379
Qy 101 AsnCysMetAenGluLeuValLeuGlnArgLysAenLeuLeuAlaGlnGluSerVal 120
Db 380 AACTGTATGACGAGCTGCTGCTCTTCAGCGGAAGAACCTGCTGGCCAGAGAGCGTG 439
Qy 121 GluThrGlnAenLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyAlaLys 140
Db 440 GAGACACAGAACTTGAAGCTGGCAGTGCATGACATGACCACCTGCAGAGCTGTACGCCAAA 499
Qy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
Db 500 CTTAAGGAGCAGTTGGAACCTGCAGCGGGGAGATGATCGGGCTTCAAGAGAGAGACAGG 559
Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAenGluLysAenGluVal 180
Db 560 CAGCTGCAGTGCAGACAGAGAGTTTGCATCAGCTCCTGAGAGATGAGAAAGATGAGGTA 619
Qy 181 GlnLysLeuGlnAenIleAlaSerArgAlaThrGlnTyrAenHisAspValLysArg 200
Db 620 CAAAAATACAAAAATATCATGCCAGCGCGGCTACTCAGTATATCATGATGTGAAGAGG 679
Qy 201 LysGluArgGluTyrAenLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db 680 AAGGAGCGTGAATATATTAAGCTTAAGGAGCGCTGCATCAGCTCGTTATGAACAAGAG 739
Qy 221 AspLysAenIleAlaMetAspValLeuAenTyrValGlyArgAlaAspGlyLysArgGly 240
Db 740 GATAAAAAACATAGCCATGGATGTTTTAAATTAATGTGGTGCAGCTGATGGCAACGAGGC 799
Qy 241 SerTrpArgThrAspLysThrGluAlaArgAenGluAspGluMetTyrLysIleLeuLeu 260
Db 800 TCATGGAGGACTGACAAAACAGAACCCAGAAATGAAGATGAGATGTACAAAATTTCTGTG 859
Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAenAlaGluLeuLysLysVal 280
Db 860 AATGATTTAGTATCCGCCAGAGACAGATCTCTGATGAGAACCGCGAGCTGAAGAAGTC 919
Qy 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
Db 920 CTCAGCAGATGAAGAAGGAGATGATCTCTCTCTCTCTCTCAGAAAGAACGCCCAGG 979
Qy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAenAspSerGly 320
Db 980 GAAAGAGCAGAGGACCGCACAGGCATGTTGCTATCTCCGATATAGAAGATGACTCTGGG 1039
Qy 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
Db 1040 GAACTGAGCAGACAGACAGCTGTGGGGCTTTCTGTGACACTGTGAGAGAGCAGCTGACA 1099
Qy 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
Db 1100 AACAGCATCAGGAAACAGTGGAGATTTTGAAAGTCATGTAGAAAAAATCGATAACCAA 1159
Qy 361 AlaSerLysValHisSerGluGlyLeuAenGluGluAspValIleSerArgGlnAspHis 380
Db 1160 GCTTCGAAGGTACACTCAGAGGGGCTTAATGAGGAGACGTCATCTCAGCAAGACCAT 1219
Qy 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400
Db 1220 GAGCAGAGACTGAGAAACTGGAGCTGGAGATTGAGCGGTGTAAAGAGATGATCAAGCT 1279
Qy 401 GlnGlnGlnLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
Db 1280 CAGCAGCAGCTCTTACAGCAGCAGCTGCGCCACCACTGTGTATGATGATGACACCACTCT 1339
Qy 421 LeuArgAspCysTyrLeuLeuLeuGluLysGluArgLeuLysGluTyrThrLeuPhe 440
Db 1340 TTGCGAGACTGTACTTGTCTGGAGAAAGAGAGCGCTTAAAGAAAGAGTGACCCCTTTT 1399
Qy 441 LysGluGlnLysLysAenPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
Db 1400 AAAGAGCAAAAAAAGAAATTTTGAGAGAGAAAGCGAAGCTTTACAGAAGCTGCCATTCGA 1459
Qy 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPhe 480
Db 1460 TTGGGTTGGAGAGAAAGCGCTTTGAAGAGAGAGCGAGCTGGGTGAAGACAGCAGCTTT 1519
Qy 481 LeuAenMetThrAenPheAspHisGlnAenSerGluAenValLysLeuPheSerAlaPhe 500
Db 1520 TTAACATGACGAACCTTTTGACCAACCACTCAGAAATGTGAAAACCTTTTCAGTGCCTTC 1579
Qy 501 SerGlySerSerAspProAspAenLeuIleValHisSerArgProArgGlnLysLysLeu 520
Db 1580 TCAGGAAGTTCTGATCCAGACATCTTATAGTCCACTCAGGCCACGCCGCAAGAGCTA 1639
Qy 521 HisSerValAlaAenGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db 1640 CACAGTGTGGCTAATGGGTGCCAGCTTGACACTCAAAAACCTGACTAAATCTCTTCTGCC 1699
Qy 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
Db 1700 TCACCTTCTACTTCAGACTTTCGCCAGACACATTCATGTGTCTGTAACACAGTTCATC 1759
Qy 561 SerValLeuAenIleThrProGluGluSerLysProSerGluValAlaIleArgGluSerThr 580
```


301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320
Db 1172 GAAAGAGCAGAGGACGGCAGCAGGACCTGTTCTATCTCCGATATAGAGATGACTCTGGG 1231
Qy 321 GluLeuSerArgAspSerValTrrGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
Db 1232 GAATGTAGCAGACAGCGGTGTGGGGCTTTCCTGTGACACTGTGAGAGAGCAGCTGACA 1291
Qy 341 AsnSerIleArgLysGlnTrrArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
Db 1292 AACAGCATCAGAAACAGTGCAGAAATTTGAAAGTTCATGTAGAAAAAATCGATAACAA 1351
Qy 361 AlaSerLysValHisSerGluClyLeuAsnGluAspValIleSerArgGlnAspHis 380
Db 1352 GCTTCGAAGTACACTCAGAGGGCTTAAATGAGAGGACGTCATCTCAGCAACAGCCAT 1411
Qy 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400
Db 1412 GAGCAAGAGACTGAGAACTGAGCTGGAGATTGAGCGGTGTAAGAGATGATCAAGGCT 1471
Qy 401 GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspThrThrSerLeu 420
Db 1472 CAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACTGTGTGATGATGACACCACCTCACTG 1531
Qy 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluTrrThrLeuPhe 440
Db 1532 TTGGCAGACTGTTACTTGTCTGGAAGAAAGCGCTTAAAGAGAGTGGACCTTTTT 1591
Qy 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
Db 1592 AAAGAGCNAAGAAATTTTTCAGAGAGAAAGGCGAAGCTTTTACAGAGCTGCCATTCGA 1651
Qy 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrrValLysGlnGlnPhe 480
Db 1652 TTGGGGTTGGAGAGAAAGCGCTTTCGAGAGAGCGAGCAGCTGGTAAAGCAGCAGTTT 1711
Qy 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
Db 1712 TTAACATGACGAACTTTGACCAACAGAACTCAGAAATGTGAAACTTTTCAGTGGCCTTC 1771
Qy 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
Db 1772 TCAGGAAGTTCTGATCCAGACAACTTATATGTCACCTACCGCCACGCGCAAGAGCTA 1831
Qy 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db 1832 CACAGTGTGGCTAATGGGTGCCAGCTTGCACATCAAAACTGACTAAATCTCTTCCTGCC 1891
Qy 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
Db 1892 TCACCTTCTACTTCAGACTTTTCGCCAGACACATTCATGTGTCTGAAACAGATTCATC 1951
Qy 561 SerValLeuAsnIleThrProGluSerLysProSerGluValAlaArgGluSerThr 580
Db 1952 AGTGTGCTGAATATAACTCTCTGAGAAAGTAAACCACTGAGGTTCGCAAGAGAAAGCAG 2011
Qy 581 AspGlnLysTrrSerValGlnSerArgProSerArgGluGlyCysTyrSerGlyCys 600
Db 2012 GATCAGAAAGTGGAGCGTGCAGTCGAGGCCAGCTCGCGGAGGGGTGTACAGCGGATGC 2071
Qy 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
Db 2072 TCCTCGGCTTTCAGGAGCGCTCAGGGGACCGAGATGATGATCTTACCT 2116
RESULT 3
BC021749
LOCUS BC021749 3425 bp mRNA linear ROD 18-JUL-2005
DEFINITION Mus musculus synovial sarcoma, X breakpoint 2 interacting protein,
mRNA (cdna clone MGC:25823 IMAGE:4165430), complete cds.
ACCESSION BC021749
VERSION BC021749.1 GI:18256805
KEYWORDS MGC.

SOURCE
ORGANISM

Mus musculus (house mouse)

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE
AUTHORS

1 (bases 1 to 3425)
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heieh, F.,
Platchenko, D., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Spapleton, T.E., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.O., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Pahey, J., Helton, E., Kettaman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Jones, S.J., Smal, U., Smal, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Mammalian Gene Collection Program Team
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26) 16999-16902 (2002)

CONSRM

TITLE

NTH MGC Project
Direct Submission
Submitted (18-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sheed, A.J., Martin, R.G., Muzny, D.M., Navevati,
A.N., Gibbs, R.A.

JOURNAL

PUBMED

REFERENCE

AUTHORS

CONSRM

JOURNAL

REMARK

COMMENT

FEATURES

source

1. 3425
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:25823 IMAGE:4165430"
/tissue_type="Salivary gland, 10 week old female mouse"
/clone_lib="NCI CGAP_SG2"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
1. 3425
/gene="Ssx2ip"
/note="synonym: Adip"
/db_xref="GeneID:99167"
/db_xref="WGI:2139150"
240..2087
/gene="Ssx2ip"
/codon_start=1
/product="synovial sarcoma, X breakpoint 2 interacting

gene

CDS

protein"
 /protein_id="AAH21749.1"
 /db_xref="GI:18256806"
 /db_xref="GeneID:99167"
 /db_xref="MGI:2139150"
 /translation="MGDMWTVDPVLCTENKNLSOYTSCTKMSPSLSYQQVLCSSVP
 LSKNVHGVFGVCTCENCTENLOSIYLDQELTTFGFSFLYEESKEAKRELINIAVLNC
 KNEELVLOKKNLLAQESVETQNLKSGDMDHLQSCYAKLKEQLETSRREMIGLOERDR
 QLOCKNRSJHQLKNEKEVEQKLQNIILASRATQYNDHVRKEREYNKLKERJHOLVMN
 KKKQNIAMDVLNVKRAQDKGSRWTKTEARNEDEMYKILLNDYERQKQTLMENAE
 LKKVLQOMKKEISLSPQKKPRERAEDGTCTVAISDIEDDSGSLSDSVMLGSCDT
 VREQLTQTKRQWRILKSHVEKLDNOAKVHSEGLNEEDVLSRQDHEOETKLELEIE
 RCKEMIKAQQQLQQLATTCDDDTTSLLRDYCLLEEKERLKEWTLKQCKKNPERE
 RRSFTAAIRLGLERKAPEERASWVKQFLNMNFDQNSENVKLFSAFSGSSDPDN
 LHVSRKQKQLHSHVANGVFACTSKLTSLPASPSSTSDFRQTHSCVSHSHSISVLNIT
 PEESKPEVARESTDQRKSVQSPRSREGYSGCSSAFSAHGDRDDLFP"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 3425
 Score: 3165.00 Matches: 615
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 Ds: 6 Gaps: 0

US-10-644-084-2 (1-615) x BC021749 (1-3425)

QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
 |||||
 Db 240 ATGGGAGATTGGATGACTGTGACACATCCAGTTCTGTGTACAGAAAACAAAATCTCTCT 299
 |||||
 QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
 |||||
 Db 300 CAATATACCTCAGAAACAAAGATGCTCCGTCCAGTTTGTACTCCACGACAGTTCTGTGC 359
 |||||
 QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrClyGlu 60
 |||||
 Db 360 TCTTCAGTACCTTTATCCAAAACAGTGATGCTGTTTTCGGTGTCTCTGACACAGGAGAG 419
 |||||
 QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
 |||||
 Db 420 AACATTGAACAAAGTATTTCTATCTTCATCAGGAGCTGACCACCTTCGGGTTCCTTCC 479
 |||||
 QY 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
 |||||
 Db 480 TTGTATGAAGAATCCAAAAGTAAAGAGCAAGAGAGAATTAATATATAGTCGCTGCTG 539
 |||||
 QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerVal 120
 |||||
 Db 540 AACTGTATGAACGAGCTGCTCGTGTCTCAGCGGAAGAACTTGCTGGCCACGAGAGCGGTG 599
 |||||
 QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnGlnSerCysTyrAlaLys 140
 |||||
 Db 600 GAGACACAGAATCTGAAGCTGGCNGTGCATGTGACACCTTCGACAGCTGCTACGCCANA 659
 |||||
 QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGlnArgAspArg 160
 |||||
 Db 660 CTTAAGGAGCAGTTGGAACCTCCAGCGCGGAGATGATCGGGCTTCAAGAGAGACAGAG 719
 |||||
 QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
 |||||
 Db 720 CAGCTGCAGTGCAAGAACAGAGAGTTTGATCAGCTCTCTGAAAGATGAGAAAGATGAGGTA 779
 |||||
 QY 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
 |||||
 Db 780 CAAAAATTACAAATATATATAGCCAGCCGGCGCTACTCAGTATATATCATGATGTGAGAGG 839
 |||||
 QY 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
 |||||
 Db 840 AAGGAGCGTGAAATATAAAGCTAAAGGAGCGCCTGCATCAGCTCGTTATGAACAAGAG 899
 |||||
 QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
 |||||

Db 900 GATAAAAAATAGCCATCGATGTTTAAATATATGTGGTTCGAGCTGATGGCAACGAGGC 959
 |||||
 QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
 |||||
 Db 960 TCATGGAGAGCTGACAAAACAGACCAGGAATGAAGATGAGATGTACAAAATCTCTGTG 1019
 |||||
 QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
 |||||
 Db 1020 AATGATTATGATACCGCCAGAGCAGATCCTGATGGAGAACGCCGGAGCTGAAGAAGGTC 1079
 |||||
 QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
 |||||
 Db 1080 CTCACGACAGATGAAGAAGAGATGATCTCTCTCTCTCTCTCAGAGAAGAACGCCAGG 1139
 |||||
 QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
 |||||
 Db 1140 GAAAGAGCAGAGGACGGCACAGGCACCTGTTCTATCTCCGATATAGAAGATGACTCTGG 1199
 |||||
 QY 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
 |||||
 Db 1200 GAACTGACGACAGACAGCGTGTGGGGCTTTCTCTGTGACACTGTGAGAGAGCAGCTGACA 1259
 |||||
 QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
 |||||
 Db 1260 AACAGCATCAGGAACACAGTGGAGAAATTTTGAAGAAGTCAATGTAGAAAAAATCGATAACCAA 1319
 |||||
 QY 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
 |||||
 Db 1320 GCTTCGAAGGTACACTCAGAGGGCTTATGAGGAGGACGTCATCTCAGCAGACAGACCAT 1379
 |||||
 QY 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400
 |||||
 Db 1380 GAGCAAGAGACTGAGAACTGGAGCTGAGAGATTGAGCGGTGTAAAGAGATGATCAAGGCT 1439
 |||||
 QY 401 GlnGlnGlnLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
 |||||
 Db 1440 CAGCAGCAGCTCTTTACAGCAGCAGCTGGCCACCACGCTGTGATGATGACACACCTCACTG 1499
 |||||
 QY 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTyrThrLeuPhe 440
 |||||
 Db 1500 TTGGCAGACTGTTTACTTGTGTAAGAAAAGAAAGCGCTTAAAGAAGAGGTGACCCCTTTT 1559
 |||||
 QY 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
 |||||
 Db 1560 AAAGCAAAAGAAAGAAATTTTGGAGAGAGAAAGGCGAAGCTTTTACAGAAGCTGCCATTCA 1619
 |||||
 QY 461 LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnGlnPhe 480
 |||||
 Db 1620 TTGGGGTTGGAGAGAAAGGGCGTTTGAAGAAGAGCGAGCGCAGCTGGGTAAAGCAGCAGTTT 1679
 |||||
 QY 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
 |||||
 Db 1680 TTAACAATGACGAATCTTTGACCACAGAACTCAGAAATGTGAACCTTTTTCAGTGTCTTTC 1739
 |||||
 QY 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
 |||||
 Db 1740 TCAGGAAGTTCTGATCTCAGACAAATCTTATAGTCCACTCACGCCACCGCAAAAGAGCTA 1799
 |||||
 QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
 |||||
 Db 1800 CACAGTGTGGCTAATGTGGGTGCCAGCTTGACATCAAAACCTGACTAAATCTCTTCTCTGCC 1859
 |||||
 QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
 |||||
 Db 1860 TCACCTTCTACTTTCAGACTTTGCCGACACACATTCATGTGTCTGAAACACAGTTCCATC 1919
 |||||
 QY 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580
 |||||
 Db 1920 AGTGTGCTGAATATAAATCTCTGGAAGAAAGTAAACCAAGTGAAGTGTGCAAGAGAAGCAGC 1979
 |||||
 QY 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCys 600
 |||||

Db 1980 GATCAGAGTGGAGCGTGCAGTCGAGGCCAGCTCGCGGGAGGGGTGCTACAGCGGATGC 2039

Qy 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
|||||

Db 2040 TCCTCGGCGCTTCAGGAGCGCTCACGGGAGCCGAGATGACTTACCT 2084

RESULT 4

BC031527 3410 bp mRNA linear ROD 29-JUN-2004

LOCUS Mus musculus synovial sarcoma, X breakpoint 2 interacting protein, mRNA (cDNA clone MGC:28268 IMAGE:4008624), complete cds.

ACCESSION BC031527

VERSION BC031527.1 GI:21594536

KEYWORDS MGC.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 3410)

AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Shat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Maruina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udén, T.B., Tohiyuki, S., Carninci, P., Prange, C., Raha, S.S., Ioquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL PROC. NATL. ACAD. SCI. U.S.A. 99 (26), 16899-16903 (2002)

PUBMED 12477932

REFERENCE 2 (bases 1 to 3410)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (08-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegod, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILLNL at: <http://image.llnl.gov>

Series: IRAP Plate: 36 Row: k Column: 10

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 20270272.

FEATURES

source Location/Qualifiers

1..3410

/organism="Mus musculus"

/mol_type="mRNA"

/strain="Czech II"

/db_xref="taxon:10090"

/clone="MGC:28268 IMAGE:4008624"

/tissue_type="Mammary tumor metastatized to lung.

MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR enhancer."

/clone_lib="NCI CGAP_Lu30"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

1..3410

/gene="Sex2ip"

/note="synonym: Adip"

/db_xref="GeneID:99167"

/db_xref="MGI:2139150"

248..2095

/gene="Sex2ip"

/codon_start=1

/product="synovial sarcoma, X breakpoint 2 interacting protein"

/protein_id="AAH31527.1"

/db_xref="GI:21594537"

/db_xref="GeneID:99167"

/db_xref="MGI:2139150"

/translation="MGDMVTVDVLTCTENKNLSQVTSKMSPPSSLYSOQVLCSSVP LSRNVHGVFGVCTGENIEQSITSLDQELTTFGPSLYEESKKEAKRELNIIVAVLC MNELLVQRKNLLAQSVEQNLKLSGDMDLQSVAKLKEQLETSRRMIGLQERDR KQDNVAMVNLVYGVADSKRGSRWRTDKTEARNEDEMYKILLNDYERQKILLMENAE LKVKQOMKEMISLSPKPKPRERABDGTGVAISDIDDSGELSRDSVWGLSCDT VREQLTNSRKQWRILSKHVKLDNQAKSVHSEGLNEEDVISRDHQEETKLELELE RSKEMIAQOQLQOQLATTCDDDTTSLRDYLLKEERLKEEWTLFKQKNFERE RRSFTEARLRLGLERAKFEERASWVKQFLLNMTFNHQNSENKILFSAFSGSDPDN LIVHSRPROKLLHSVANGVPACTSKLTSLPASPTSDPROTHSCVSEHSISVLNIT PEESKPSVARESTDQKMSVQSPSSREGCYSCSSAFRSRHAHGRDRLP"

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	3410
Score:	3162.00	Matches:	614
Percent Similarity:	100.0%	Conservative:	1
Best Local Similarity:	99.8%	Mismatches:	0
Query Match:	99.9%	Indels:	0
DB:	6	Gaps:	0

US-10-644-084-2 (1-615) x BC031527 (1-3410)

Qy	1	MetGlyAspTrpMetThrValThrAppProValLeuCysThrGluAonLysAsnLeuSer	20
Db	248	ATGGGAGATTGGATGACTGTGCAGATCCAGTTCGTGTACAGAAAACAAAATCTCTCT	307
Qy	21	GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys	40
Db	308	CAATATACCTCAGAAACAAAGATGTCTCCGTCAGTTGTACTCCCAAGCAAGTCTGTGC	367
Qy	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu	60
Db	368	TCITTCAGTACCTTTATCCAGAAATGTGATGGTGTTCCTCGGTGTCTTCGCACAGGAG	427
Qy	61	AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer	80
Db	428	AACATTGAAACAAAGATATTTCTTATCTTGATCAGAGCTGACCACTTCGGGTTCCTTCC	487
Qy	81	LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAenlleValAlaValLeu	100
Db	488	TTGTATGAAGANTCCAAAAGTAAAGAGGCAAGAGAGAAATTAATATAGTCGTGTTCTG	547
Qy	101	AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerVal	120
Db	548	AACCTGATGAACAGAGTGTCTGCTGCTTCAGCGGAAGAACCTGTGTCGCCACGAGAGCG	607
Qy	121	GluThrClnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys	140
Db	608	GAGACACAGAACTTGAAGCTGGGGCAGTGACATGACCACCTGCAGAGCTGCTACGCCAAA	667

QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
 DB CTTAAGAGGAGGTGTGAACAGTCCAGCGGAGAGATGATCGGCTTCAAGAGAGACAGG 727
 QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180
 DB CAGCTGCAGTGCAGAACAGGAGTTTGATCAGCTCCTGAGAGATGAGAAAGATGAGTGA 787
 QY 181 GlnLysLeuGlnAsnIlelelaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
 DB CAAAATTTACAAATATCATAGCAGCGGGTACTCAGTATATCATGATGTAAGAGG 847
 QY 201 LysGluArgGluTyrAsnLysLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
 DB AAGGAGCGGAGATATAAAGCTAAAGGAGCGCCTGCATCAGCTCGTTATGAAACAGAG 907
 QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
 DB GATAAAACATAGCCATGATGTTTAAATTTATGTGGTCGAGCTGATGGCAACAGGC 967
 QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
 DB TCATGGAGGACTGACAAAACAGAGCCAGGAATGAAGATGAGATGTACAAAATTCGTGTG 1027
 QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysVal 280
 DB AATGATTATGATGACCGCAGAGAGAGATCCTGATGTGAAGCCGCGAGCTGAAAGAGGTC 1087
 QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
 DB CTCAGCAGATGAGAGAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1147
 QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320
 DB GAAAGAGCAGAGAGCGGACAGGACTGTTGCTATCTCGATATAGAAGATGACTCTGGG 1207
 QY 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluLeuThr 340
 DB GAACGTGAGCAGAGACAGCGGTGGGGCCCTTCTCTGTGACACTGTGAGAGAGCAGCTGACA 1267
 QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
 DB AACACATCAGGAACACAGTGGAGATTTTGAAGTCAITGTAAGAAATCGATNAACCA 1327
 QY 361 AlaSerLysValHisSerGluGlyLeuAsnGluAspValIleSerArgGlnAspHis 380
 DB GCTTCGAAGGTGCACTCAGAGGGCCTTAATGAGGAGGAGCGTCTCTCAGCAACAGACCAT 1387
 QY 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400
 DB GAGCAAGAGACTGAGAAACTGGAGCTGGAGATTGAGCGGTGTAAAGAAATGATCAAGGCT 1447
 QY 401 GlnGlnGlnLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
 DB CAGCAGCAGCTCTTACACAGCAGCTGCCACCGCTGTGTGATGATGACACACCTCACTG 1507
 QY 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLysGluGluTrpThrLeuPhe 440
 DB TTGCAGACTGTACTTACTTCTGGAAGAAAGGAAAGCGCCTTAAAGAGAGGTGGACCCCTTTT 1567
 QY 441 LysGluGlnLysLysAsnPheGluArgGluArgArgSerPheThrGluAlaIleArg 460
 DB AAAGAGCAAAAGAAAGAAATTTTGAAGAGAAAGGCGGAGCTTTTACAGAAAGCTGCCATTCGA 1627
 QY 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPhe 480
 DB TTGGGGTTGGAGAGAAAGCGGTTTGAAGAGAGCGGAGCGGCTGGTGAAGAGCAGGATTT 1687
 QY 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
 DB TTAACATGACGAACCTTTTGACCACCAGAACTCAGAAATGTGAAACTTTTTCAGTGCCTTC 1747

QY 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
 DB TCAGGAAGTCTTGATCCAGACAATCTTATAGTCCACTCACGCCACGGCAAGAAGCTA 1807
 QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
 DB CACAGTGTGGCTAATGGGGTGCAGCTTGACATCAAACTGCTAAATCTCTTCTCTGCC 1867
 QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
 DB TCACCTTCTTACTTCAGACTTTTCGCCAGACACATTCATGTGTCTGTGACACAGTTCATC 1927
 QY 561 SerValLeuAsnIleThrProGluLysSerLysProSerGluValAlaArgGluSerThr 580
 DB AGTGTGCTGAATATTAATCTCTCAAGAAAGTAAACCAAGTGAAGTTCGAAGAGAAAGCAG 1987
 QY 581 AspGlnLysTyrSerValGlnSerArgProSerArgGluGlyCysTyrSerGlyCys 600
 DB GATCAGAAGTGGAGGCTGAGTCCGAGGCCAGCTCGCGGAGGGGTGTCTACAGTGTGATGT 2047
 QY 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615
 DB TCCTCGGCTTCAGGAGCGCTCACGGGACCGAGATGACTTACCT 2092
 RESULT 5
 LOCUS BC078687
 DEFINITION Rattus norvegicus synovial sarcoma, X breakpoint 2 interacting protein, mRNA (cdna clone MGC:93056 IMAGE:7115887), complete cds.
 ACCESSION BC078687
 VERSION BC078687.1 GI:50926757
 KEYWORDS MGC.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 2547)
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smallos, D.E., Schnerch, A., Schein, J.B., Jones, S.J. and Marra, M.A.
 Mammalian Gene Collection Program Team
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 2 (bases 1 to 2547)
 JOURNAL
 PUBMED 12477932
 REFERENCE
 AUTHORS
 CONSRM
 TITLE
 Mammalian Gene Collection Program Team
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 2 (bases 1 to 2547)
 DIRECT Submission
 NIH MGC Project
 Submitted (02-AUG-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: http://mgc.nci.nih.gov
 Contact: MGC help desk
 Email: mgc@nci.nih.gov
 Tissue Procurement: Howard Jacobs
 cDNA Library Preparation: Express Genomics
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdpaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>
 Series: IRAK Plate: 183 Row: b Column: 1
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 31342223.

FEATURES

source	1..2547	Location/Qualifiers
	/organism="Rattus norvegicus"	
	/mol_type="mRNA"	
	/db_xref="taxon:10116"	
	/clone="MGC:93056 IMAGE:7115887"	
	/tissue_type="Testis, rat (Brown Norway)"	
	/clone_lib="NIH MGC_237"	
	/lab_host="DH10B"	
	/note="Vector: pExpress1"	
gene	1..2547	
	/gene="Ssx2ip"	
	/note="synonym: ADIP"	
	/db_xref="GeneID:308023"	
	/db_xref="RGD:708490"	
CDS	366..2207	
	/gene="Ssx2ip"	
	/codon_start=1	
	/product="synovial sarcoma, X breakpoint 2 interacting protein"	
	/protein_id="AAH78687.1"	
	/db_xref="GI:50926758"	
	/db_xref="GeneID:308023"	
	/db_xref="RGD:708490"	
	/translation="MGDMVTVDPLVCTENKLSQYTSSTKMSPLSYSOQVLCSTP LSKNVGVFSAGTGENITEOSLYLQDELTTFGPSLYESKSKEARELSIVALNLC MQLLVLRKNLLAQESVETONKLGSDMHLQSCYAKLKEQLEASRREMISIQERD KKDNIAMDVLNYGVGRGKRSWRDTKTEARNEDEMYKILLNDYEQKQILLENAB LKVLQOQKEMISLLSPQKKPREAEDSTGVTVSDVEDDAGELSRDGVMSLSCDT RCKEMIAQOQLQALACDDDTTSLLRDQYLLEKEREKEMSLFKQKKNFEBE RRSTEAAIRLGERKAFEEERASWVKQFLNMTTFDQNSENVKLFSAFSGSDPDN LIVPRFRQKPHSVANGVPACTSKLAKSLPTSFSDFPSRSCVSEHSPVSALITVPE ETKPNVEGRESTDQKWSVVRSPSREGCYGCGSSAYTSSHVERDDL"	
ORIGIN		
Alignment Scores:		
Pred. No.:	9,62e-301	Length: 2547
Score:	2937.00	Matches: 567
Percent Similarity:	95.1%	Conservative: 18
Best Local Similarity:	92.2%	Mismatches: 28
Query Match:	92.8%	Indels: 2
DB:	6	Gaps: 1
US-10-644-084-2 (1-615) x BC078687 (1-2547)		
Qy	1	MetGlyAspThrPmetThrValThrAspProValLeuCystrGluAsnLysAsnLeuSer 20
Db	366	ATGGAGATTGGATGACCTTTACAGATCCAGTCTGTGTACAGAAACAAAAATCTCTCT 425
Qy	21	GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db	426	CAATATACCTCAGAAACAAAGATGTCCTCGTCAAGTTTATATCTCGCAGCAAGTACTGTGC 485
Qy	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db	486	TCTGCAACACCTTTATCCAAAGATGTGCATGGTGTTCAGTGCCTCTGCACAGGAGAG 545
Qy	61	AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrPheGlyPheProSer 80
Db	1626	TTGCGAGACTGTTACTTGTCTTGAAGAAAGGAAACGCCCTTAAAGAAAGAGTGGTCCCTTTT 1685
Db	546	AACATCGAACAGAGTATTTTCGTATCTTGATCAGAACTGACTACCTTCGCTTCCTCCCTTC 605
Qy	81	LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db	606	TTGTATGAGAAGTCCAAAAGTAAAGGAGGCGAAGCGAGAGTAAAGTATAGTGTCTTCTGT 665
Qy	101	AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerVal 120
Db	666	AATCGCATGAATGAGCTGCTTGTCTTACGCGAAGAACCTCTCGGCCCAAGAAGAGGTG 725
Qy	121	GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
Db	726	GAGACACAGAACTCTGAAGCTGGCAGTGACATGACACCACTGTCAGAGCTGCTACGCCAAA 785
Qy	141	LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
Db	786	CTTAAAGCAACAGTTGGAGGCTCCAGCGAGAGATGATCAGCCTTTCAGSAGAGACAGAGA 845
Qy	161	GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180
Db	846	CAGCTACAGTGCAAAACAGGAATTTGTCATCAGCTCTCTGAANAACGAAAGAAAGAGGTA 905
Qy	181	GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Db	906	CAAAAATTTACAAAATATCATAGCCAGTCGGGCTACTCAGTATATCATGATGTGAAGAGA 965
Qy	201	LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db	966	AAGGAGCGGAGTACATAAATGAAGGAGCGCTGTCATCAGCTGTGTATGAACAAGAG 1025
Qy	221	AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db	1026	GATAAAAATATAGCCATGGACGTTTAAATTAAGTGGCCGAGTGGAGGAGGAGGC 1085
Qy	241	SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
Db	1086	TCCTGGAGGACTGATAAAAACAGAAAGCCAGGAATGAAGATGAAATGTACAAAATTTCTGCTG 1145
Qy	261	AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
Db	1146	AATGATTATAGTACCCAGAGAGAGATCCTCTCTGAGAGTGGGAGCTGGAAGAGCTC 1205
Qy	281	LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
Db	1206	CTCCAGCAGATGAAGAAAGAGATGATCTCTCTCTTCTCTCAAAAGAAAGAAACCCAGA 1265
Qy	301	GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
Db	1266	GAAGAGCAGAGGACAGCAGGCACTGTTGTCTATCTCCGATGTAGAAGACGACGCTGGG 1325
Qy	321	GluLeuSerArgAspSerValTyrGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
Db	1326	GAGCTGAGCAGAGATGGTGTGGAGCCTTTCTCTGTGACACTGTTCAGGAGCAGCTTACA 1385
Qy	341	AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
Db	1386	AACAGCATCAGGAAGCAGTGGAGATTCGAAAGCCATGTGAAAGCACTGTGAAACCA 1445
Qy	361	AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
Db	1446	GCTTCAAGGTACACTCAGAGGCTTTCTTGAAGAGAGACGTCATCTCAGCAAGACCAT 1505
Qy	381	GluGlnGluThrGluLysLeuGluLeuIleLysArgCysLysGluMetIleLysAla 400
Db	1506	GAGCAGAGAGACTGAGAAACTGGAGCTGGAGATTGAGCGGTGTAAAGAGATGATCAAGGCT 1565
Qy	401	GlnGlnGlnLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
Db	1566	CAGCAGCAGCTCTTACAGCAACAGCTGGCCACTGCGTGTGATGACGACACACCTCACTG 1625
Qy	421	LeuArgAspCysTyrLeuLeuGluLysGluArgLeuLysGluGluTyrThrLeuPhe 440
Db	1626	TTGCGAGACTGTTACTTGTCTTGAAGAAAGGAAACGCCCTTAAAGAAAGAGTGGTCCCTTTT 1685

```

QY      441  LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaIleArg 460
Db      1686  AAGAGCAAAAGAAATTTTGGAGAGAGAACGACGCTTTACAGAGCTCTATTCCG 1745

QY      461  LeuGlyLeuGluArgLysAlaPheGluGluArgGluAlaSerTrpValLysGlnGlnPhe 480
Db      1746  TTGGGGTGGAGAGAAAGCGTTTGGAGAGAGCAGCAGCTGGTGAAGCAGCAGTTT 1805

QY      481  LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
Db      1806  TTAACATGACGACGCTTTGATCACAGAACTCAGAAAATGTGAAACTTTTCAGTGCCTT 1865

QY      501  SerGlySerSerAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
Db      1866  TCAGGAAGTTCTGATCCAGACAACTTATAGTCCACCCACGCGCCACGCAAAAGAGCCA 1925

QY      521  HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db      1926  CACAGTGTCTNATGGGTACACGCTGCACATCAAACTGGCTAAGTCTCTTCCGACC 1985

QY      541  SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
Db      1986  TCACCT-----TCAGACTTCTGCCGCTCTCGCTCATGTGTCTGTGAGCACAGTCCCGTC 2039

QY      561  SerValLeuAsnIleThrProGluGlnSerLysProSerGluValAlaArgGluSerThr 580
Db      2040  AGTGGCGTGACTGTGACTCTCTGAAGAAACCAACCGAATGAGGTTTGGAGAGAAAGTACG 2099

QY      581  AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGlyCys 600
Db      2100  GACCAAGAGTGGAGCGTGTGTCCAGACCCAGCTCCCGGAGGGTGTCTACGGTGGATGC 2159

QY      601  SerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615
Db      2160  TCTCGGCTACACAAGTCCCACTCGGACGAGATGACTTACCA 2204

RESULT 6
AF532970
LOCUS   Rattus norvegicus afadin- and alpha-actinin-binding protein ADIP
DEFINITION
ACCESSION AF532970
VERSION   AF532970.1 GI:27451840
KEYWORDS
SOURCE    Rattus norvegicus (Norway rat)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 3195)
          Asada,M., Irie,K., Morimoto,K., Yamada,A., Ikeda,W., Takeuchi,M.
          and Takai,Y.
          ADIP, a Novel Afadin- and alpha-Actinin-Binding Protein Localized
          at Cell-Cell Adherens Junctions
          J. Biol. Chem. 278 (6), 4103-4111 (2003)
          12446711
REFERENCE 2 (bases 1 to 3195)
          Asada,M., Irie,K. and Takai,Y.
          Direct Submission
          Submitted (26-JUL-2002) Department of Molecular Biology and
          Biochemistry, Osaka University Graduate School of Medicine/Faculty
          of Medicine, Yamada-oka 2-2, Suita 565-0871, Japan
          Location/Qualifiers
FEATURES             source
     1..3195
     /organism="Rattus norvegicus"
     /mol_type="mRNA"
     /strain="Sprague-Dawley"
     /db_xref="taxon:10116"
     79..1920
     /codon_start=1
     /product="afadin- and alpha-actinin-binding protein ADIP"
     /protein_id="AAO15016.1"

```

```

/db_xref="GI:27451841"
/translation="MGDMVTVDPVLCITENKNLSQYTSGETKMSPSLSYQQVLCSATP
LSKNVGVFSAPCTGENIEQSYLDQDELTTFFPFSLYEESKEAKELSGIVALLNC
NELLVLQKNLLAQESVETQNLKLGSDMDHLQCYAKLKEQLSEARREMSLQERDR
QLOCKNRNLHOLLNKEEVEQKQNI TASRATQYNHDKRKEREYVKLKERLHOLVMN
KKDKNAMDVLNVYGRVDKSGSWRTOKTEARNEDEMYKIILNDYEVYKQKILLNNAE
LKKVLQOMKEMISLLSPQKKPRERAEDSTGTVVVISDVEDDAGELSGRDGSKLSCDT
VREGUTSIRKQWRLKSHVEKLDNQASKVHSGFHEBDVISRDHEQETBKLEUEIE
RCKEMIKAQQLQQLATACDDDTTSLLRDCLLLEKERLKEMSLFPKQKNPFERE
RRSTFAAQRLLERKAFEBERASWVQQLFNMNTTFHQNSENVKLSFAFGSSDPDN
LIVHPRPROKKPHSVANGVPACTSKLAKSLPTSPDFCPSRSCVSEHSPVSALTVTPE
ETKENVREBSTDQKNSVVRPSSREGYCGCSSAYTSSHVERDDLP"

```

ORIGIN

```

Alignment Scores:
Pred. No.:      1,32e-300      Length:      3195
Score:          2937.00        Matches:    567
Percent Similarity: 95.1%      Conservative: 18
Best Local Similarity: 92.4%    Mismatches: 28
Query Match:    92.8%          Indels:      2
DB:             6              Gaps:       1

US-10-644-084-2 (1-615) x AF532970 (1-3195)

QY      1  MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
Db      79  ATGGGAGATTGGATGACTGTTTACAGATCCAGTCTCTGTGTACAGAAAAACAAAATCTCTCT 138

QY      21  GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db      139  CAATATACCTCAGAAAAACAAGATGTCTCCGTCAGTTTATATCTCGACAAGTACTGTGTC 198

QY      41  SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db      199  TCTGCAACACCTTTATCCAGAAATGTCATCGTGTTCAGTGCCTTCGCAGGAGAGAG 258

QY      61  AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrPheGlyPheProSer 80
Db      259  AACATCGAACAGAGATATTTCTGTATCTTATCAGGAACTGACTACCTTCGGTTTCCTTCC 318

QY      81  LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db      319  TTGTATGAAGAATCCAAAAGTAAGAGCGAGAGGAGATTAAAGTATAGTTAGTTCTCTCTG 378

QY      101  AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
Db      379  AACTGCATGATGAGCTGCTTGTCTTCAGCGGAGAAACCTCTCTGGCCACGAGAAAGCGTG 438

QY      121  GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
Db      439  GAGACACAGAATCTGAAGCTGGGCGAGTGACATGACACACACTGCAGAGCTGTACGCCAAA 498

QY      141  LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGlnArgAspArg 160
Db      499  CTTAAGGAACAGTGTGGAGGCTTCAGCGAGAGATGATCAGCCTTCAGGAGAGAGACAGA 558

QY      161  GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
Db      559  CAGCTACAGTGCAAAAACAGGAATTTGCATCAGCTCTCTGAAAACAGGAAGAGAGGTA 618

QY      181  GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Db      619  CAAAATTTACAAAATATATAGCCAGTCCAGTGGGCTACTCAGTATATATCATGATGTGAAGA 678

QY      201  LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db      679  AAGGAGCGGAGTAGTACAATAAATGAAGAGCGCTCTGCATCAGCTTGTATTATGAACAAGAG 738

QY      221  AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db      739  GATAAAAATATAGCCATGGAGCTTTTAAATTTACGTGGCGCGAGTGGATGGAAACGAGGCG 798

QY      241  SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260

```

	799	TCTCGGAGGACTGATAAAACAGAAGCCAGGAATGAAGATGTAACAAAATTCTGCTG	858
Db			
Qy	261	AenAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAenAlaGluLeuLysLysVal	280
Db			
Db	859	AATGATTATGATACCCGCCAGNACAGATCCTGCTGGAGATCGGAGCTGTAGNAGGTC	918
Qy	281	LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg	300
Db			
Db	919	CTCCAGCAGATGAAGAAAGAGATGATCTCTCTCTCTCTCAAAAGAAACCACCA	978
Qy	301	GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly	320
Db			
Db	979	GAAGACGAGAGSACACACAGGCACCTGTGTCTCATCTCCGATGTAGAGACGAGCTGGG	1038
Qy	321	GluLeuSerArgAspSerValTIpGlyLeuSerCysAspThrValArgGluGlnLeuThr	340
Db			
Db	1039	GAGCTGAGCAGATGGTGTTGGAGCCTTTCCTGTGACACTGTTCAGGAGCAGCTTACA	1098
Qy	341	AmnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln	360
Db			
Db	1099	AACAGCATCAGAAAGCAGTGAGAAATCTGAAAAACCATGTGAAAAAATTGATAACCAA	1158
Qy	361	AlaSerLysValHisSerGluGlyLeuAsnGluAspValIleSerArgGlnAspHis	380
Db			
Db	1159	GCATTCAAGGTACACTCAGAGGGCTTTTCATGAAGAGGACGTCACTCACACACAGACCAT	1218
Qy	381	GluGlnGluThrGluLysLeuGluGluIleGluArgCysLysGluMetIleLysAla	400
Db			
Db	1219	GAGCAAGAGACTGANAACCTGGAGCTGGAGATTGAGCGGTGAAGAGATGATCAAGCT	1278
Qy	401	GlnGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspThrThrSerLeu	420
Db			
Db	1279	CAGCAGCAGCTCTTACAGCAACAGCTGGCCACTCGGTGTGATGACGACACCCACTCACTG	1338
Qy	421	LeuArgAspCysTyrLeuLeuGluGluLysGluArgIleLysGluGluTrpThrLeuPhe	440
Db			
Db	1339	TTGGCAGACTGTACTTCTGTTGAAGAAAAAGAACGCCCTTAAAGAAAGAGTAGTGCTCCCTTTTT	1398
Qy	441	LysGluGlnLysLysAenPheGluArgGluArgArgPheThrGluAlaAlaIleArg	460
Db			
Db	1399	AAAGAGCAAAAAAGAAATTTTGAGAGAGAAAGACGACTTTACAGAAAGCTGTATTTCGC	1458
Qy	461	LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPhe	480
Db			
Db	1459	TTGGGTTGGAGAGAAAGCCGTTTGAGGAAGAGCGAGCCAGCTGGGTGAAGCAGCAGTTT	1518
Qy	481	LeuAenMetThrAenPheAspHisGlnAenSerGluAenValLysLeuPheSerAlaPhe	500
Db			
Db	1519	TTAAACATGACGACCTTTGATCACCAGAACTCAGAAAAATGTGAAACTTTTTCAGTGCCTTT	1578
Qy	501	SerGlySerSerAspProAspAenLeuIleValHisSerArgProArgGlnLysLysLeu	520
Db			
Db	1579	TCAGGAAGTTCTGATCCAGACAACTTATATGTCACCCCAGCCGCCGCAAAAGAGCCCA	1638
Qy	521	HisSerValAlaAenGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla	540
Db			
Db	1639	CACAGTGTCTGCTAATGGGTTGCCAGCTTGACATCCAACTGGCTAAGTCTCTTCCGACC	1698
Qy	541	SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle	560
Db			
Db	1699	TCACCT-----TCAGACTTCTGCCCGTCTCGCTCATGTGTGTCTGAGCACAGTCCCGCTC	1752
Qy	561	SerValLeuAenIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr	580
Db			
Db	1753	AGTCGGCTGACTGTGACTCTCTGAAAGAACCCAACCGAATGAGGTGGAAGAGAAAGTACG	1812
Qy	581	AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCys	600
Db			
Db	1813	GACCAGAACTGAGCGTGGTGTCCAGACCCAGCTCCCCGGAGGGTGTCTACGGTGGATGC	1872
Qy	601	SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro	615
Db			

Db 1873 TCTCGCGCTACACAGCTCCCGGACGAGATCACTTACCA 1917

RESULT 7
BC033637
LOCUS
DEFINITION
BC033637 mRNA linear PRI 27-JAN-2004
Homo sapiens synovial sarcoma, X breakpoint 2 interacting protein,
mRNA (cDNA clone MGC:45118 IMAGE:5578582), complete cds.
BC033637
BC033637.1 GI:21707136
MGC.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 3220)
Strausberg R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Jordan, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Zordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Utsid, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullany, S.J., Boeak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Pahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,
Schnerfeld, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 3220)
Strausberg, R.
Direct Submission
Submitted (02-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NIHC)
Gaithersburg, Maryland
Web site: <http://www.nisic.nih.gov/>
Contact: nisic_mgc@nih.gov
Blakesley, R.W., Bouffard, G.G., Stren, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Teague, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 69 Row: 0 Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7662381.
Location/Qualifiers
1. .3220
/organism="Homo sapiens"

```

/dbl_type="mRNA"
/dbl_xref="taxon:9606"
/clone="MGC:45118 IMAGE:5578582"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH MGC_67"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..3220
/gene="SSX2IP"
/note="SSX2IP"
/note="synonyms: KIAA0923, ADIP"
/db_xref="GeneID:117178"
152..1996
/gene="SSX2IP"
/codon_start=1
/product="SSX2IP protein"
/protein_id="AAH33637.1"
/db_xref="GI:21707137"
/db_xref="GeneID:117178"
/translation="MGDMVTDPGLSSSEKTIISQYTSKTPSPSLYSQQVLCCSIP
LSKNVHSPFSAFCTEDNIEQISLYLDQBLTTFGFPSPLYEESKGTRELNIVAVLNC
MNELLYLQKLNLAQENVETQNLKLSDMHQLQSCYSKLQEOLETSRREMIGLOQRDR
QLQCKRNHLQKLNKNEDEVQKLNIIASRATQYNDHMKRERENKLRHLQVLWN
KDKKIAMDLNIGRADRGKSGWRTGTEARNEDEMYKILINDYREYKQILMENAE
LKKVLQMKKEMISLLSPKKPRPRVDSTGTVISDVEEDAGELSRSEMDLSCETV
REOLNLSRKQWRILKSHVEKLDNOVSKVHLEGFNDEDI SRODHEORTKLELEIQ
CKEMIKTQQQLLQQLATAYDDDTLSLLRDCYLLBEKRLKEKSLFKBQKNFERER
RSFTAAALRLGLERKAFERASRLWKQFLNMVTFDQNSENVKLFSPAGSSDMDNL
IVHSQPKQKPHSVNSGSPVCMKSLTKSLPASPSDFCQTRSCISEHSSINVLNITA
BEIKPNQVGGECTNQKWSVASRPGSQEGCYSGLSYTNSHVEKDLIP"
ORIGIN
Alignment Scores:
Pred. No.: 1 256-284 Length: 3220
Score: 2786.50 Matches: 538
Percent Similarity: 93.2% Conservative: 35
Best Local Similarity: 87.5% Mismatches: 41
Query Match: 88.0% Indels: 1
DB: 5 Gaps: 1
US-10-644-084-2 (1-615) x BC033637 (1-3220)
Qy 1 MetGlyAspTyrMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
Db 152 ATGGAGATTGGATGACTTTACAGATCCAGGTCTGTCTTCAGAAAGCAAAATCTATCTCT 211
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db 212 CAATATACCTCAGAAACAAAGATGCTCCATCAAGTTTATATCTACACCAAGTGTATGT 271
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 272 TCCTCAATACCTTTATCGAAAATGTGCACAGTTTCTTCAGTGCCTTCTGCACAGAAGAT 331
Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGlnLeuThrThrPheGlyPheProSer 80
Db 332 AATATTGACAGAGTATCTATCTGATCAGGAATGACTACTTTTGGTTTTCCTTCA 391
Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db 392 TTATATGAAGAAATCCAAAGGTAAGAGACAAAGAGAGAGATTTAAATATAGTAGTGTACTA 451
Qy 101 AsnCysMetAsnGluLeuValLeuValLeuGlnArgLysAsnLeuAlaGlnGluSerVal 120
Db 452 AATTGTATGAATAGCTGCTTGTGCTTCAGGGGAGAACCTTCTAGCTTCAGGAAAATGTG 511
Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
Db 512 GAGACACAGAAATTTGAAGCTGGGAAGTATGATGACCAATCTACAGAGCTGCTACTCAAA 571
Qy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
Db 572 CTTAAGGAACAACCTGGAAACCTCCAGGAGGGGAAATGATTGGCTTCAGGAAGAGACAGA 631

```



```
Db      190 ATGGGAGATTGGATGACTGTTACAGATCCAGGTCTGTCTTCAGAAAGCAAAACTATCTCT 249
Qy      21  GlnTyrThrSerGluThrLysMetSerProSerLeuTyrSerGlnValLeuCys  40
Db      250 CAATATACCTCAGAAACAAGATGCTCCATCAAGTTTATATCTCACAGCAAGTGTATGT 309
Qy      41  SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu  60
Db      310 TCTTCAATACCTTTATCGAAANNTGTGCACAGCTTTTTCAGTGCCTTCTGCACAGAAGAT 369
Qy      61  AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer  80
Db      370 AATATTGAACAGAGTATCTCATATCTTGATCAGGAATTGACTACTTTTGGTTTTCCTTCA 429
Qy      81  LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu  100
Db      430 TTATATGAAGAAATCCAAAGGTAAGAGACAAGAGAGAGTTAAATATATAGTAGCTGTACTA 489
Qy     101  AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal  120
Db      490 AATTGTATGAATGAGCTGCTTGTGCTTCAGCGGAAGAACCTTCTAGCTCAGGAANAATGTG 549
Qy     121  GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnInsrCysTyrAlaLys  140
Db      550 GAGACACAGAATTTGAACCTCGGAAGTGAATGACCATCTACAGAGCTGTACTACAAA 609
Qy     141  LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg  160
Db      610 CTTAAGGAAACAACCTCGAAACCTCCAGGAGGAAATGATTGGCTTCAGGAAGAGACAGA 669
Qy     161  GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal  180
Db      670 CAGTTTACAATGTAGAAACAGGAATTTGTCATCAGCTACTTAAAGAAATGAGAAAGATGAGGTG 729
Qy     181  GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg  200
Db      730 CAAAAATTACAAATATCATTTGCAAGTCGAGCTACTCAGTATATATCATGATATGAGAGA 789
Qy     201  LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys  220
Db      790 AAAGAGCGTGAAATATAATAAACTGAAGCAAGCTCTACATCAACTTGTGTATGAACAAGAAA 849
Qy     221  AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly  240
Db      850 GATGAAGAAAATAGCTATGGAATTTTGAATTTATGCCGGGAGAGCTGATGGAAGAAAGAGGC 909
Qy     241  SerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu  260
Db      910 TCCTTGGAGGACTGGTAAAACCTGAAGCCAGGAAATGAAGATGAATGTATAAGATTCTCTTG 969
Qy     261  AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal  280
Db      970 AATGATTATGAATATCGTCAGAAAACAAATCCTAATGGAATAATGCAGAACTTTAAGAGGTT 1029
Qy     281  LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg  300
Db     1030 CTTCAACAAATGAAAAAGGAATGATTCTCTCTTCTTCTCCCCCAAAAAGAAAAACCTAGA 1089
Qy     301  GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly  320
Db     1090 GAAAGAGTAGATAGTACAGGAACCTGTT---ATTTCGATGTTGAAGAAAGATGCCGGG 1146
Qy     321  GluLeuSerArgAspSerValTyrGlyLeuSerCysAspThrValArgGluGlnLeuThr  340
Db     1147 GAACTAAGCAGAGAGTAGTATGGGACCTTTCTCTGTGAAACCTGTGAGAGAGCAGCTTACA 1206
Qy     341  AsnSerIleArgLysGlnTyrArgIleLeuLysSerHisValGluLysLeuAspAsnGln  360
Db     1207 AACAGCATCAGAAAACAGTGGAGAAATTTTGAAGAGTCATGTAGAAAAAGCTTTGATAACCAA 1266
Qy     361  AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis  380
Db     1267 GTTTCAAAGGTACACCTGGGAAGGTTTTAATGATGAGATGTAATCTCCAGGCAAGACCAT 1326
```

```
Qy      381  GluGlnGluThrGluLysLeuGluLeuGluArgCysGlyMetIleLysAla  400
Db     1327 GAACAAGAAACTGAAAAAATCGAGTTAGAAATTCAGCAGTGTAAAGAAATGATTAAAACT 1386
Qy      401  GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu  420
Db     1387 CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGTCATATGATGATGATACCACTTCACCTA 1446
Qy      421  LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTyrThrLeuPhe  440
Db     1447 TTACGAGACTGTTATTTGTTGGAAGAAAGGAAAGCTCTCAAGAAAGAAAGTGGTCCCTTTT 1506
Qy      441  LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg  460
Db     1507 AAAGAGCAGAAAAAGAAATTTTGAAGAGGAGAGACGAGCTTTACAGAAAGCGCTATTGCG 1566
Qy      461  LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTyrValLysGlnGlnPhe  480
Db     1567 CTGGGATTTGGAGAGAAAGGCAATTTGAAGAAAGAAAGAGCCAGCTGGTTAAAGCAGCAGTTT 1626
Qy      481  LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe  500
Db     1627 CTAAATATGACTACTCTTTGACCACGAGAACTCAGAAAATGTGAAAATCTTTTCAGTCCCTTC 1686
Qy      501  SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu  520
Db     1687 TCAGAAAGTTCTGATTTGGGACAACTTTATAGTCACCTCGAGCGCCGCAAAAAGAGCCT 1746
Qy      521  HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla  540
Db     1747 CACAGTGTGTCTAATGGGCTCTCCAGTTTGCATGTCTAAACTTACTAAATCTCTTCTCTGCT 1806
Qy      541  SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle  560
Db     1807 TCACCTTCCACTTCAGACTTTTGCACAGACAGCTTCTGTCATATCTGAACATAGTTCAATC 1866
Qy      561  SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaAArgGluSerThr  580
Db     1867 AATGTACTGAATATAACTGCTGGAAGAAATTAACCAAAATCAGGTGGAGTAGAATGTACA 1926
Qy      581  AspGlnLysTyrSerValGlnInsrArgProSerSerArgGluGlyCysTyrSerGlyCys  600
Db     1927 AATCAAAAATGGAGTGGGCATCAAGACCTTGATTCACAGGAAGGTTGCTATAGTGGATGC 1986
Qy      601  SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro  615
Db     1987 TCCTTGAGCTACACAAATTTCTCATGTAGAAAAGATGACTTACCT 2031
```

RESULT 11

```
AR718997 LOCUS 2272 bp DNA linear PAT 07-OCT-2005
DEFINITION Sequence 1497 from patent US 6943241.
ACCESSION AR718997
VERSION AR718997.1 GI:77367951
KEYWORDS
SOURCE Unknown.
```

ORGANISM

Unknown.

REFERENCE

```
1 (bases 1 to 2272)
Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,
Yamanoto, J.-i., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,
Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and
Masubo, Y.
```

TITLE

Full-length cDNA

JOURNAL

Patent: US 6943241-A 1497 13-SEP-2005;

Research Association for Biotechnology; Tokyo;

FEATURES

source

1. .2272

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores:				
Pred. No.:	8.66e-278	Length:	2272	
Score:	2720.00	Matches:	536	
Percent Similarity:	86.8%	Conservative:	34	
Best Local Similarity:	81.6%	Mismatches:	40	
Query Match:	85.9%	Indels:	47	
DB:	2	Gaps:	3	
US-10-644-084-2 (1-615) x AR718997 (1-2272)				
Qy	1	MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer	20	
Db	224	ATGGAGATTGGATGACTGTTACAGATCCA-----GAAACCAAACTATCTCT	271	
Qy	21	GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys	40	
Db	272	CAATATACCTCAGAAACAAAGATGCTCCATCAAGTTTATACCTCACAGCAAGTGTATGT	331	
Qy	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu	60	
Db	332	TCATTCAATACCTTTATCGAAAAATGTGCACAGTTTTTCAGTGCCTTCTGCACAGAAGAT	391	
Qy	61	AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer	80	
Db	392	AATATTGACAGAGTACTCATATCTTGATCAGGAATGTACTACTTTTGGTTTCTCTTCA	451	
Qy	81	LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu	100	
Db	452	TTATATGAAGAAATCCAAAGGTAAAGACAAAGAGAGAGTAAATATAGTAGCTGTACTA	511	
Qy	101	AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnLeuSerVal	120	
Db	512	AATTGTATGAATAGCTGCTTGTGCTTCAGCGGAAGAACCTTCAGCTCAGGAATGTG	571	
Qy	121	GluThrGlnLeuLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys	140	
Db	572	GAGACACAGAATTTGAACGGGAAGTGATATGACCATCTACAGAGCTGCTACTCAAA	631	
Qy	141	LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg	160	
Db	632	CTTAAGGAACAACCTGGAACCTCCAGAGGGAAATGATTTGGCTTCAGGAAGAGACAGA	691	
Qy	161	GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal	180	
Db	692	CAGTTACAATGTAAAGACAGGAATTTGCTCAGCTACTAAAGAAATGAGAAAGATGAGGTG	751	
Qy	181	GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg	200	
Db	752	CAAAAATTACAAATATCATTTGCAAGTCGACTACTCAGTATATCATGATATGACAGAGA	811	
Qy	201	LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys	220	
Db	812	AAAGAGCGTGAATATATAAATCAAGGAACCTCTACATCAACTTGTATGAACAAGAAA	871	
Qy	221	AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly	240	
Db	872	GATTAAGAAAAATAGCTATGACATTTTGAATTTATGTCGGGAGAGCTGATGAAAAAGAGGC	931	
Qy	241	SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu	260	
Db	932	TCCTGGAGGACTGGTAAACTGAACCCGAGGAAATGAAATGATATAAATTTCTCTTG	991	
Qy	261	AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal	280	
Db	992	AATGATTATGAATATCGTCAGAAAACAAATCTTAATGAAAAATGCAAACTTAAGAGGTT	1051	
Qy	281	LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg	300	
Db	1052	CTTCAACAAATGAAAAAGGAATGATTTCTTCTTCTCCCAAAAAAGAAACCTTAGA	1111	
Qy	301	GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly	320	

Db	1112	GAAAGAGTAGATGATAGTACAGGAACCTGTT--ATTTCGATGTTGAAGAAGATGCCGGG	1168
Qy	321	GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr	340
Db	1169	GAACTAAGCAGACAGATATGTGGACCTTTCTGTGTGAACCTGTGAGAGAGCTTACA	1228
Qy	341	AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln	360
Db	1229	AACAGCATCAGAAACAGTGGAGANTTTTGAAGAAGTCATGTAGAAAGCTTGTATACCAA	1288
Qy	361	AlaSerLysValHisSerGluGlyLeuAsnGluAspValIleSerArgGlnAspHis	380
Db	1289	GTTTCAAGGTACACCTGGAAAGGTTTAAATGATGAAGATGTAATCTCACGACAAGCAT	1348
Qy	381	GluGlnGluThrGluLysLeuGluLeuIleGluArgCysLysGluMetIleLysAla	400
Db	1349	GAAACAGAACTGAAAAAATCGAGTTAGAAATTCAGCAGGTGTAAGAAATGATTAATACT	1408
Qy	401	GlnGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu	420
Db	1409	CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCATATGATGATGATACACTTCACTA	1468
Qy	421	LeuArgAspCysTyrLeuLeuGluLysGluArgLeuLysGluGluTyrThrLeuPhe	440
Db	1469	TTACGAGACTGTTATTTTGTGGAAAGAAAGAACGTCCTCAAGAAAGAAATGTCCTTTT	1528
Qy	441	LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg	460
Db	1529	AAAGACAGAAAAAGAAATTTTGAGAGGAGAGACGAAGCTTTACAGAGCGCTATTTCGC	1588
Qy	461	LeuGlyLeuGlu-----	464
Db	1589	CTGGGATTTGGAGATTGGCTTTCTCTGCCAAGTGTTTTACACAAGAAATCTCTTTGGGT	1648
Qy	464	-----	464
Db	1649	ATCCATCATTCATCCACGGTCAGTAGGTGTGTGTGTCAGAAAGACAGTGCACCAAGACTAT	1708
Qy	465	-----ArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnG	479
Db	1709	ATTGTGCTGTTAAAAAAGAAAGGCATTTTGAAGAAGAAAGAGCCAGTGGTTAAAGCAGCA	1768
Qy	479	nPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAl	499
Db	1769	GTTTCTTAATATGACTACCTTTGACCACACAGAACTCAGAAATGTGAAACTTTTCAGTGC	1828
Qy	499	aPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLy	519
Db	1829	CTTCTCAGGAAGTCTCTGATTGGGACAATCTTATAGTGCACTCGAGCGCAGCGCAAGAA	1888
Qy	519	sLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuPr	539
Db	1889	GCCTCAGTGTGCTAAATGGGTCTCCAGTTTCCATGTCTAAACTTACTAAATCTCTTCC	1948
Qy	539	oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe	559
Db	1949	TGCTTACCTTCCACTTCAGACTTTTGCACGACACCGTTCCTGCTATCTCAACATAGTTC	2008
Qy	559	rIleSerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSe	579
Db	2009	AATCAATGACTCAATATAACTGCTGAAGAAATTAACCAAAATCAGGTGGAGGAGAACG	2068
Qy	579	rThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGl	599
Db	2069	TACAAATCAAAATGAGTGTGGCGTCAAGACCTGGATCATCAGGAAGGTGCTATAGTGG	2128
Qy	599	YCysSerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro	615
Db	2129	ATGCTCTCTGAGCTACACAAATTTCTCATGTAGAAAAAGATGACTTACCT	2177

RESULT 12
AX747972
LOCUS

AX747972 2272 bp mRNA linear PAT 20-JUN-2003

DEFINITION Sequence 1497 from Patent EP1308459.
ACCESSION AX747972 GI:32132360
VERSION AX747972.1
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1
REFERENCE
AUTHORS Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Negahari,K. and
Masuho,Y.
TITLE Full-length cDNA sequences
JOURNAL Patent: EP 1308459-A 1497 07-MAY-2003;
Helix Research Institute (JP) ; Research Association for
Biotechnology (JP)
FEATURES
source
Location/Qualifiers
1..2272
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 8,666-278 Length: 2272
Score: 2720.00 Matches: 536
Percent Similarity: 86.8% Conservative: 34
Best Local Similarity: 81.6% Mismatches: 40
Query Match: 85.9% Indels: 47
DB: 2 Gaps: 3
US-10-644-084-2 (1-615) x AX747972 (1-2272)
Oy 1 MetGlyAspTTPMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
Db 224 ATGGAGATTGGATGACCTTTACATCCA-----GAAAGCAAACATCTCTCT 271
Oy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db 272 CAATATACCTCAGAAACAAAGATGCTCCATCAAGTTTATCTACACGCAAGTGTATGT 331
Oy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 332 TCTTCAATACCTTTATCGAAAAAATGTGCACAGTCTTTTTCAGTGCCTTCTGCACAGAAGAT 391
Oy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
Db 392 AATATTGACACAGAGTATCTCATATCTTGATCAGGAATTGACTCTTTTGGTTTCCCTTCA 451
Oy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db 452 TTATATGAAGAATCCAAAGGTAAAGACACAAAGAGAGAGATTAAATATAGTAGTGTACTA 511
Oy 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerVal 120
Db 512 AATTGTATGAATAGCTCTGTGTCTTCAGCGGAAGAACCTTCTAGCTTCAGGAATAATGTG 571
Oy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
Db 572 GAGACACAGAATTTGAACCGGAAGTATGATGGACCATCTACAGAGCTGCTACTCAAAA 631
Oy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGlnLysAspArg 160
Db 632 CTTAAGGAACAACCTGGAACCTCCAGAGAGGGAATGATTGGGCTTCAGGAAGAGACAGA 691
Oy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
Db 692 CAGTTACAATGTAAGAACAGGAATTTGCATCAGCTACTTAAAGAAATGAGAAAGATGAGGTG 751
Oy 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200

Db 752 CAAAAAATTACAAAATATATCAATTGCCAAGTCGAGCTACTCAGTATAAATCATGATATGAAGAGA 811
Oy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db 812 AAAGAGCGTGAATATAATAAACTGAAGGAACTGTACATCACTTGTATGAACAAGAAA 871
Oy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db 872 GATAGAAAAATAGCTATGGACATTTTGAATTTATGTCCGGAGAGCTGATGTAAGAAAGAGGC 931
Oy 241 SerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
Db 932 TCCTCGAGAGACTGGTAAACCTGAAGCCAGGAATGAAGATGAATGTATATAAAATCTCTTTG 991
Oy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
Db 992 AATGATTATGAATATCGTCAGAAACAATCCTAATGGAATATGCAGAACTTAAAGAAGTT 1051
Oy 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
Db 1052 CTTCAACAATGAAAGAGAAATGATTCTCTCTCTCTCCCAAAAGAAAGAACTAGTA 1111
Oy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320
Db 1112 GAAAGAGTAGATAGTAGTACAGGAACCTGTT---ATTTCCGATGTTGAAGAAATCCCGG 1168
Oy 321 GluLeuSerArgAspSerValTyrPglyLeuSerCysAspThrValArgGluGlnLeuThr 340
Db 1169 GAACCTAACGACAGAGAGATGATGGGACCTTTCTGTGAACCTGTGAGAGAGCAGCTTACA 1228
Oy 341 AsnSerIleArgLysGlnTyrArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
Db 1229 AACGACATCAGAAAAACAGTGGAGAAATTTGAAAGTCTATGTAGAAAAAGCTGATACCAA 1288
Oy 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
Db 1289 GTTTCAAAGGTACACCTCGAAGGTTTTAATGATGAAGATGTAATCTCAGCAACAGACCAT 1348
Oy 381 GluGlnGluThrGluLysLeuGluLeuGluArgCysLysGluMetIleLysAla 400
Db 1349 GAACAAGAACTGAAAAACCTGAGTTAGAAATTCAGCAGTGTAAAGAAATGATTAATAACT 1408
Oy 401 GlnGlnGlnLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
Db 1409 CAGCAACAGCTTTTACAGCAGCAGCTGCTACTGCATATGATGATGATACCACTTCACTA 1468
Oy 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTyrThrLeuPhe 440
Db 1469 TTACGAGACTGTTATTTGTTGGAAGAAAGAAAGCGTCTCAAAGAAAGAAATGTCCTCTTT 1528
Oy 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
Db 1529 AAAGAGCAGAAAAAGAAATTTGAGAGGAGAGAGAGAGCTTTTACAGAAGCCGCTATTGCG 1588
Oy 461 LeuGlyLeuGlu----- 464
Db 1589 CTGGGATTGGAGATTGGCTTTCTCCTGCCAAGTGTGTTTACACAAGAAATCTCTTTTGGGT 1648
Oy 464 ----- 464
Db 1649 ATCCATCATTCACCGGTACAGTAGGTGCTCAGAGACAGACGTGCAACCAAGACTAT 1708
Oy 465 -----ArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnG 479
Db 1709 ATTTGCTGTTTAAAAAAGAAAGCAATTTGAAGAAAGAAAGAGCCAGTGGTTAAAGCAGCA 1768
Oy 479 nPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAl 499
Db 1769 GTTTCTAAATATGACTACCTTTGACCACCAAGAACTCAGAAAAATGTGAAACTTTTTCAGTGC 1828
Oy 499 aPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLys 519
Db 1829 CTTCTAGGAAGTTCTGATTGGGCAATCTTATAGTGCACCTCGAGGCGCAGCGCAAGAA 1888

Qy 519 sLeuHisSerValalabenglyValProAlaCvThrSerlystleuThrvSerleuPr 539
 Db 1889 GCCTCACAGTGTCTAATGGGTCCTCAGTTTGATGCTAAACTTACTAAATCTCTTCC 1948
 Qy 539 oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe 559
 Db 1949 TGCCTTCCACTTCCAGACTTTTCCGACACACGTTCCGTGCATATCTGAACATATCTC 2008
 Qy 559 rLleSerValleuAenlleThrProGluGluSerLyseProSerGluValalalaArgGluSe 579
 Db 2009 AATCAATGTACTGAATATACTGCTGAAGAATAATTAACCAAAATCAGGTGTGGAGAGAAGC 2068
 Qy 579 rThrAspGlnlysrProSerValGlnSerArgProSerSerArgGluGlyCvThrvSerGl 599
 Db 2069 TACAAATCAAAATGGAGTGTGGCGTCAAGACCTTGGATCAGAGGAAGTGTCTATATGG 2128
 Qy 599 YCysSerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
 Db 2129 ATGCTCTTGAGCTACACAAATCTCATGTAGAAAAGATGACTTACT 2177
 RESULT 13
 AK093173
 LOCUS Homo sapiens cDNA FLJ35854 f1s, clone TESTI2007113. 2272 bp mRNA linear PRI 20-JAN-2006
 DEFINITION
 AK093173
 ACCESSION
 AK093173.1 GI:21751964
 VERSION
 keywords oligo capping; f1s (full insert sequence).
 SOURCE
 Homo sapiens
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.
 REFERENCE
 AUTHORS
 Oca, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,
 Wakamatsu, A., Hayaishi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,
 Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, K.,
 Iehi, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,
 Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,
 Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,
 Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furiya, T.,
 Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,
 Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,
 Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,
 Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,
 Yoshida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi, F., Hara, R.,
 Hara, H., Tanase, T. O., Nomura, Y., Togiya, S., Komai, F., Hara, R.,
 Takeuchi, K., Arita, M., Imose, N., Mueashino, K., Yuuki, H., Oshima, A.,
 Sabaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,
 Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Sato, N., Takami, S.,
 Teraehima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,
 Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,
 Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,
 Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,
 Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,
 Ozaki, K., Hirao, M., Omori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,
 Inagaki, H., Ikeda, Y., Okamoto, S., Okitani, R., Kawakami, T.,
 Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,
 Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,
 Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,
 Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,
 Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,
 Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.
 Complete sequencing and characterization of 21,243 full-length
 human cDNAs
 Nat. Genet. 36 (1), 40-45 (2004)
 14702039
 2
 Iehibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,
 Ono, Y., Hotuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,
 Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,
 Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,
 Wakamatsu, A., Iehi, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y.,

Saito, K., Nishikawa, T., Kimura, K., Mateuo, K., Nakamura, Y.,
 Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A.,
 Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
 Nagahari, K., Masuho, Y., Nagai, K., and Isogai, T.
 NEDO human cDNA sequencing project
 Unpublished
 3 (bases 1 to 2272)
 Isogai, T. and Yamamoto, J.
 Direct Submission
 Submitted (04-JUL-2002) Takao Isogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail: flj-cdna@nifty.com, Tel:81-438-52-3975, Fax:81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction: Helix Research Institute (HRI) (supported by Japan
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
 HRI, and Biotechnology Center, National Institute of Technology and
 Evaluation; clone selection for full insert sequencing: HRI and
 RAB; annotation: HRI and RAB.
 FEATURES
 Location/Qualifiers
 1..2272
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="TESTI2007113"
 /tissue_type="testis"
 /clone_lib="TESTI2"
 /note="cloning vector: pME18SFL3"
 ORIGIN
 Alignment Scores:
 Pred. No.: 8 66e-278 Length: 2272
 Score: 2720.00 Matches: 536
 Percent Similarity: 86.8% Conservative: 34
 Best Local Similarity: 81.6% Mismatches: 40
 Query Match: 85.9% Indels: 47
 DB: 3 Gaps: 3
 US-10-644-084-2 (1-615) x AK093173 (1-2272)
 Qy 1 MetGlyAspTrpMetThrValThrAppProValLeuCysThrGluAsnLysAsnLeuSer 20
 Db 224 ATGGGAGATTGGATGACTGTTCACAGATCCA-----GAAAGCAAAACTATCTCT 271
 Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
 Db 272 CAATATACCTCAGAAAACAAAGATGCTCCATCAAGTTTATCTACACAGCAAGTGTCTATGT 331
 Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
 Db 332 TCTTCAATACCTTTATCGAAAANAATGTGCACAGTTTTTTCAGTGCCITTCGCACAGAAGAT 391
 Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
 Db 392 AATATTGAACACAGATGATCTCATATCTTGTATCAGAAATGACTACTTTTGGTTTTTCTCTCA 451
 Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
 Db 452 TTATATGAAGAATAATCCAAAGTAAGAGACAAAGAGAGAGTTAATATAGTACTGTACTA 511
 Qy 101 AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
 Db 512 AATTGTATGAATGAGTGTCTTGTGCTTCAGCGGAAGAACCTTCTAGCTCAGGAAAATGTG 571
 Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
 Db 572 GAGACACAGAAATTTGAAGCGGGAAGTATGATATGACCATCTACAGAGCTCTCTACTCAAAA 631
 Qy 141 LeuLysGluGlnLeuGluThrSerArgGluMetIleGlyLeuGlnGluArgAspArg 160
 Db 632 CTTAAGGAACAACTGGAAAACCTCCAGAGGGGAATGATGTTGGCTTTCAGGAAGAAGACAGA 691

Qy	161	GlnLeuGlnCysLysAsnArgSerLeuthisGlnLeuLeuLysAsnGluLyAspGluVal	180
Db	692	CAGTTTACAATGTGAAGAACAGGAATTTGCATCAGCTACTATAAAGAATAGCAAAAGATGAGCGTG	751
Qy	181	GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyvAssnHisAspVallysArg	200
Db	752	CAAAAATTACAAAAATCATTTGCCAAGTCGAGCTACTCAGTATATAATCATGATATGAAGAGA	811
Qy	201	LysGluArGGlutyvAsnLysLeuLysGluArgLeuHissGlnLeuValMetAsnLysLys	220
Db	812	AAAGAGCGTGAATAATAATAAACCTGAAGGAACGCTACATCAACTGTGTTATGAACAAGAA	871
Qy	221	AspLysAsnIleAlaMetAspValLeuAsnTyvValGlyArGAlaAspGlyLyArGgly	240
Db	872	GATAAGAAAAANTAGCTATCGACATTTTGAATTTATGTCGGGAGAGCTGATGGAAGAAAGAGCC	931
Qy	241	SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyvLysIleLeuLeu	260
Db	932	TCCTGGAGGACTGGTAAACCTGAAGCCAGGAATGAGATGAATGTTATPAAATTTCTCTTGT	999
Qy	261	AsnAspTyvGluTyvArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal	280
Db	992	AATGATTATGAATATCGTCAGAAACAAATCCTAATGGAAATGCAGAACTTAAGAGGTT	1051
Qy	281	LeuGlnGlnMetLysLysGluMetLleserLeuLeuSerProGlnLysLysLysProArg	300
Db	1052	CTTCAACAATGAAAGAGGAATGATTTCTCTCTCTCTCCCCAAAAGAAGAAACCTAGA	1111
Qy	301	GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly	320
Db	1112	GAAGAAGTAGATGATAGTACAGGAACCTGTT---ATTTCGAGATGTTGAAGAAGATGCGCGG	1168
Qy	321	GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr	340
Db	1169	GAACTAAGCACAGAGAGTATGTGGGACCCTTTCCTGTGNAACCTGTGAGAGACGACCTTACA	1228
Qy	341	AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln	360
Db	1229	AACAGCATCAGAAAACACGTGAGAAATTTTGAAGAGTCATGTAGAAAAAGCTTGATAACCAA	1288
Qy	361	AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis	380
Db	1289	GTTTTCAAGGTACACCTGGGAAGGTTTAAATGATGAAGATGTPAATCTCAGCACAAGACCAT	1348
Qy	381	GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla	400
Db	1349	GAACAAGAACTGAAAAACCTCGAGTTAGAAANTCAGCAGGTGAAGAAATGATTAANAAT	1408
Qy	401	GlnGlnGlnLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu	420
Db	1409	CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCATATGATGATGATGCCACTTCACTA	1468
Qy	421	LeuArgAspCysTyvLeuLeuGluGluLysGluArgLeuLysGluGluTrpThrLeuPhe	440
Db	1469	TTACGAGACTGTTATTTGTTGGAAGAAAAGGAAGCGTCTCAAGAAAGAAATGGTCCCTTTTT	1528
Qy	441	LysGluGlnLysLysAsnPheGluArgGluArgArgSerPheThrGluAlaAlaIleArg	460
Db	1529	AAAGACGAGAAAAGAAATTTTGAGAGGGAGNAGCAAGCTTTACAGAAAGCGGCTATTTCGC	1588
Qy	461	LeuGlyLeuGlu-----	464
Db	1589	CTGGGATTGGAGATTGGCTTCTCCTGCCAAGTGTTTTTACACAAGAAATCTCTTTTGGGT	1648
Qy	464	-----	464
Db	1649	ATCATCATTCATCCACGGTCAGTAGGTGTGGTGTGAGAGACAGTGCACCAAGACTAT	1708
Qy	465	-----ArgLysAlaPheGluGluGluArgAlaSerTrpVallysGlnGl	479
Db	1709	ATTGTGCTTTAAAAAGAAAGGCAATTTGAAGAAAGAAAGGCCAGTGTGGTTTAAAGACGAC	1768
Qy	479	nPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAl	499

Db	1769	GTCTTAAATGACCTACCTTTGACCAACAGAACTCAGAAAATGTGAAACTTTTCAGTGC	1828
Qy	499	aPeSerGlySerSerAspProAspAnLeuIleValHisSerArgProArgGlnLysLy	519
Db	1829	CTTCTCAGGAAGTTCGATTGGGACAACTTATAGTGCACCTCGAGCGACGCCGCAAGAA	1888
Qy	519	sLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuPr	539
Db	1889	GCCTCAGAGTGTCTAATGGGTCTCCAGTTTGCATGCTCTAACTTACTTAATCTCTTCC	1948
Qy	539	oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe	559
Db	1949	TGCTTCACTTCCACTTCAGACTTTTGGCCAGACACAGTTCCTGCATATCTGAACATAGTTC	2008
Qy	559	rIleSerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSe	579
Db	2009	AATCAATGTACTGAATATAACTCTCTGAAGAAATTAACCAAAATCAGCTTGGAGGAGAAGC	2068
Qy	579	rThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGl	599
Db	2069	TACAAATCAAAATGAGTGTGGCGTCAAGACCTGGATACAGGAAGTTGCTATAGTGG	2128
Qy	599	yCysSerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro	615
Db	2129	ATGCTCTTGGAGCTACACAAATCTCATGTAGAAAAAGATGACTTACCT	2177
RESULT 14			
BC064389			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT			

Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mc@nih.gov
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
 Maduro, O.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripoop, S., Thomas, P.J., Touchman, J.W.,
 Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 139 Row: k Column: 13
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 41281570.

FEATURES

source

1..2540
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:75026 IMAGE:5745142"
 /tissue_type="Brain, adult medulla"
 /clone_lib="NIH MGC_119"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 1..2540

gene

/gene="SSX2IP"
 /note="synonyms: KIAA0923, ADIP"
 /db_xref="GeneID:117178"
 /db_xref="MIM:608690"

CDS

241..1920
 /gene="SSX2IP"
 /codon_start=1
 /product="SSX2IP protein"
 /protein_id="AAH64389.1"
 /db_xref="GI:39963670"
 /db_xref="GeneID:117178"
 /db_xref="MIM:608690"
 /translation="MGDMVTDPGLSSSEKTSIQVTSFTKSPSSLYSQVLCSSIP
 LSKNVHFFSAFCTEDINIQSYLDQELTFPGPSLYESKGETKRELNIVAVLNC
 MNELVLRKNLQAENVETQNLKIGSDMDHLQSCYSLKEQLETSRREMGLEQDRD
 QLOCKRNLHOLLKNEDEVOKLQNIASRATQYNHDMKEREYNKLERLHOLVNN
 KDKKIAMDILNVYGRADKGGKSWRTKTEARNEDVMYKILLNDYEVROKQILWENAE
 LKKVLOMKKEMISSLSPQKKPRVDDSGTVISDVEADGELSRESMDLSCETV
 REQNTSIRKQWRILKSHVEKLDNQSKVHLGENDSDVSRQHEQETKELEIQQ
 CKEMIKTQQLQOLATAYDDDTSLLRDCYLLDEERLKEEWSLFKEQKNFERRR
 SFTBAALRLGLERKAFEEERASWLKQDFLNMTTFHQNSENVKLFSAFSGSSDMDNLI
 VHSRQPKQKPHSVNGSPVCMKSLTKSLPASPTSDFCQTRSCISEHRAKN"

ORIGIN

Alignment Scores:
 Pred. No.: 1,21e-274 Length: 2540
 Score: 2691.00 Matches: 536
 Percent Similarity: 80.6% Conservative: 35
 Best Local Similarity: 75.7% Mismatches: 42
 Query Match: 85.0% Indels: 96
 DB: 5 Gaps: 3

US-10-644-084-2 (1-615) x BC064389 (1-2540)

Qy 1 MetGlyAspThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
 Db 241 ATGGAGATTGGATGACTGTTACAGATCCAGGTCTGTCTTCAGAAAGCAAACTATCTCT 300
 Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnValLeuCys 40

Db 301 CAATATACCTCAGAAAAAAGAGATGCTCCATCAAGTTTATATACACAGCAAGTCTATGT 360
 Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
 Db 361 TCTTCAATACCTTTATCGAAAAATGTCACAGTTTTTTTCAGTGCCTTCTCGACAGAAGAT 420
 Qy 61 AsnIleGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
 Db 421 AATATTGAACAGAGATATCTCATATCTTGATCAGGAATTGACTACTTTTGGTTTCTTCA 480
 Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
 Db 481 TTATATGAGAATCCAAAGGTAAGAGACAAAGAGAGAGTTAAATATATAGTAGCTGTA 540
 Qy 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
 Db 541 AATTGATGATGAGTGTCTTGCTTCACGGAGGAACCTTCTAGCTCAGGAATAATGTG 600
 Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
 Db 601 GAGACACAGANTTTGAAGCTGGGAAGTGATATCGACCATCTACAGAGCTCTACTCAAAA 660
 Qy 141 LeuLysGluGlnLeuGluThrSerArgGluMetIleGlyLeuGlnGluArgAspArg 160
 Db 661 CTTAAGGAACAACCTCGGAACCTCCAGGAGGGAATGATTGGCTTCAGGAAGAGACAGA 720
 Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180
 Db 721 CAGTTCAATGTGAAGAACAGGAATTTGCATCAGCTACTAAGAATGAGAAGATGAGGTG 780
 Qy 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
 Db 781 CAAAAATTACAAATATCATTTGCAAGTCGAGCTACTCAGTATATCATGATATGAAGAGA 840
 Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
 Db 841 AAAGAGCGTGAATATATAAATGAACTGAAGAACGCTACATCAACTGTTGTTATGAACA 900
 Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
 Db 901 GATAAGAAATAGCTATGGACATTTTGAATTATGTCGGGAGAGCTGATGGAAGAGAGGC 960
 Qy 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
 Db 961 TCCTGGAGGACTGCTAAACCTGAAGCCAGGAATGAAAGTGAATGTAATAAATCTCTTG 1020
 Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
 Db 1021 AATGATTTATGAATATCGTCAGAAACAAATCCTAATGAAATGCAGAACTTAAGAAGGTT 1080
 Qy 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
 Db 1081 CTTCAACAATGAAAAGGAAATGATTTCTCTTCTTCTCCCAAGAAAGAAACCTAGA 1140
 Qy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
 Db 1141 GAAAGAGTAGATGATAGTACAGGAACCTGTTT--ATTTCCGATGTTGAAGAAGATGCCGG 1197
 Qy 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
 Db 1198 GAACTAAGCAGAGAGAGATGTCGGGACCTTTCTCTGTGAACTGTGAGAGAGCAGCTTACA 1257
 Qy 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
 Db 1258 AACAGCATCAAAAACAGTGGAGAAATTTTGAAGAATCATGTAGAAAAGCTTGTATACCAA 1317
 Qy 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
 Db 1318 GTTTCAAAGGTACACCTCGGAAGGTTTTTAATGATGAAGATGTAATCTCACGACAGACCAT 1377
 Qy 381 GluGlnGluThrGluLysLeuGluLeuGluArgCysLysGluMetIleLysAla 400
 Db 1378 GAAACAAGAAACTGAAAACCTCGAGTTAGAAATTCAGCAGCTGTAAAGAAATGATTAAACT 1437


```
Db 426 CAGTATAATCATGATATGAAGAGAAAAGAGCGTGAATATAATTAACCTGAAGACGCTCTA 485
Qy 213 HisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrVal 232
|||
Db 486 CATCAACTGTTTATGAAACAAGAAAGATAAGAAAATAGCTATGGACATTTTGAATTATGTC 545
Qy 233 GIVArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaAaArgAsnGlu 252
|||
Db 546 GGGAGAGCTGATGGAAAAGAGGCTCTCTGGAGGACTGGTAAACTGAAGCCAGCAAGTAA 605
Qy 253 AspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMet 272
|||
Db 606 GATCAATGTATATAAAATCTCTTCAATGATATATCAATATCGTCAGAAAACAAATCCTAATG 665
Qy 273 GluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeu 292
|||
Db 666 GAAAATGCAGAACTTAAGAGAGTTCTTCAACAAATGAAAAGGAATGATTTCTCTCTT 725
Qy 293 SerProGlnLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIle 312
|||
Db 726 TCTCCCAAAAAGAAAGAACCTAGAGAAAGAGTAGATGATAGTACAGGAACGTGT---ATT 782
Qy 313 SerAspIleGluAspAspSerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCys 332
|||
Db 783 TCCGATGTTGAAGAAGATGCCGGGNACTAAGCAGAGAGAGTATGTGGACCTTTTCCTGT 842
Qy 333 AspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSer 352
|||
Db 843 GAAACTGTGAGAGAGCAGCTTTACAAACAGCATCAGAAAACAGATGGGAGAAATTTTGAAAGT 902
Qy 353 HisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGluGlyLeuAsnGluGlu 372
|||
Db 903 CATGTAGAAAAGCTTGATAACCAAGTTTCAAGGTACACCTGGGAAGGTTTTTAATGATGAA 962
Qy 373 AspValIleSerArgGlnAspHisGluGlnGluThrGluLysLeuGluLeuGluIleGlu 392
|||
Db 963 GATGTAATCTCACGACAGACCATGNAACAAGAAACTGAAAACTCGAGTTAGAAAATTCAG 1022
Qy 393 ArgCysLysGluMetIleLysAlaGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThr 412
|||
Db 1023 CAGTGTAAAGAAATGATTAATAACTCAGCAACAGCTTTT---CAGCAGCTCGCTACTGCA 1079
Qy 413 CysAspAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLysGluArg 432
|||
Db 1080 TATGATGATGATACACCTTCACTATTACGACCTGTTATTGTTGGAGAGAAAAGAACGT 1139
Qy 433 LeuLysGluGluTyrThrLeuPheLysGluGlnLysLysAsnPheGluArgGluArg 452
|||
Db 1140 CTCAAAGAGAAATGGTCCCTTTTAAAGAGCAGAAAAGAAATTTTGAGGGGAGAGACGA 1199
Qy 453 SerPheThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArg 472
|||
Db 1200 AGCTTTACAGAAAGCCGCTATTTCGCTGGGATTGGAGAGAAAGGCATTTGAAAGAGAAAGA 1259
Qy 473 AlaSerTrpValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGlu 492
|||
Db 1260 CCCAGTTGGTTAAGCAGCAGTTTCTAATAATAGTACTCTTTGACCACCAAGACTCAGAA 1319
Qy 493 AsnValLysLeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHis 512
|||
Db 1320 AATGTGAAACTTTTCAGTGCCTTCTCAGGAAGTTCTGATTGGGACAAATCTTATAGTGCAC 1379
Qy 513 SerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCysThrSer 532
|||
Db 1380 TCGAGGCCAGCCGCAAAAGAACCTCACAGTGTGTCTAATGGGTCTCCAGTTTGCATGTCT 1439
Qy 533 LysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArgGlnThrHisSer 552
|||
Db 1440 AAACTTACTAATACTCTCTCTGCTTCCACTTCCAGACTTTTGGCCAGACACGTTCC 1499
Qy 553 CysValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGluSerLysPro 572
|||
Db 1500 TGCATATCTGAACATAGTTCAATCAATGTACTGAATATAACTGCTGAAGAAATTTAAACCA 1559
```

```
Qy 573 SerGluValAlaArgGluSerThrAspGlnLysTrpSerValGlnSerArgProSerSer 592
|||
Db 1560 AATCAGGTTGGAGGAGAAACGTACAAATCAAAAATGGAGTGTGGCTCAAGACCTGGATCA 1619
Qy 593 ArgGluGlyCysTyrSerGlyCysSerSerAlaPheArgSerAlaHisGlyAspArgAsp 612
|||
Db 1620 CAGGAAGG-TGCTATAGTGGATGCTCTCTTGAGCTACACAAAATTTCTCATGTAGAAAAAGAT 1678
Qy 613 AspLeuPro 615
|||
Db 1679 GACTTACCT 1687
```

Search completed: June 14, 2006, 01:46:27
Job time : 8263 secs

GenCore version 5.1.1.9
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2006, 20:22:20 ; Search time 14775 Seconds
(without alignments)
11651.198 Million cell updates/sec

Title: US-10-644-084-1
Perfect score: 2692
Sequence: 1 cgtaggagagtgcacaggagc.....ttgattatggagtaataatgggg 2692

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_env.*
- 2: gb_pat.*
- 3: gb_ph.*
- 4: gb_pl.*
- 5: gb_pt.*
- 6: gb_ro.*
- 7: gb_sta.*
- 8: gb_sy.*
- 9: gb_un.*
- 10: gb_vi.*
- 11: gb_ov.*
- 12: gb_htg.*
- 13: gb_in.*
- 14: gb_om.*
- 15: gb_ba.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2692	100.0	2692	6 AF532969	Mus muscu
2	2692	100.0	3424	6 AK173064	Mus muscu
3	2688.8	99.9	3425	6 BC021749	Mus muscu
4	2676	99.4	3410	6 BC031527	Mus muscu
5	2189.4	81.3	3195	6 AF532970	Rattus no
6	1810.2	67.2	2547	6 BC078687	Rattus no
7	1701.4	63.2	3220	5 BC033637	Homo sapi
8	1701.4	63.2	5835	5 AB023140	Sequence
9	1699.8	63.1	5835	5 AB023140	Sequence
10	1512.2	56.2	3073	5 AY651262	Homo sapi
11	1492.6	55.4	2379	5 AY367055	Homo sapi
12	1290.6	47.9	2540	5 BC064389	Homo sapi
13	1260.4	46.8	2716	2 BD156617	Primer fo
14	1260.4	46.8	2716	2 AX877359	Sequence
15	1260.4	46.8	2716	5 AK001710	Homo sapi
16	1258.8	46.8	2272	2 AR718997	Sequence
17	1258.8	46.8	2272	2 AX747972	Sequence
18	1258.8	46.8	2272	5 AK093173	Homo sapi

19	944	35.1	165330	6	AC124987	Mus muscu
c 20	895.2	33.3	185929	12	AC017085	AC017085 Homo sapi
c 21	895.2	33.3	188840	5	AC046144	AC046144 Homo sapi
c 22	698.2	25.9	255776	12	AC142185	AC142185 Rattus no
c 23	698.2	25.9	306870	12	AC098557	AC098557 Rattus no
c 24	589	25.6	233457	12	AC106173	AC106173 Rattus no
25	560	20.8	826	2	BD148918	BD148918 Primer fo
26	560	20.8	826	2	AX868856	AX868856 Sequence
27	559.8	20.8	3892	11	BC070749	BC070749 Xenopus 1
28	547.8	20.3	2594	11	BC060934	BC060934 Danio rer
29	530.8	19.7	3218	11	BC070587	BC070587 Xenopus 1
30	421.8	15.7	781	11	EX934353	EX934353 Gallus ga
c 31	399.8	14.9	161874	5	AC114482	AC114482 Homo sapi
c 32	399.8	14.9	179844	5	AC009265	AC009265 Homo sapi
33	338.8	12.6	1503	2	BD158988	BD158988 Primer fo
34	338.8	12.6	1503	2	AX881380	AX881380 Sequence
35	338.8	12.6	1503	5	AK021515	AK021515 Homo sapi
c 36	334.2	12.4	563	2	BD150903	BD150903 Primer fo
c 37	334.2	12.4	563	2	AX870841	AX870841 Sequence
38	323	12.0	498	2	BD058793	BD058793 Sequence
39	312.2	11.6	674	11	CR353768	CR353768 Gallus ga
40	305.8	11.4	403	2	AX071986	AX071986 Sequence
41	297.2	11.0	428	2	BD026751	BD026751 Sequence
42	297.2	11.0	428	2	AR727012	AR727012 Sequence
43	297.2	11.0	428	2	AX887141	AX887141 Sequence
44	218.2	8.1	258641	12	AC127120	AC127120 Rattus no
45	212	7.9	884	2	BD148462	BD148462 Primer fo

ALIGNMENTS

RESULT 1
AF532969
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS

AF532969 2692 bp mRNA linear ROD 03-FEB-2003
Mus musculus afadin- and alpha-actinin-binding protein ADIP mRNA,
complete cds.
AF532969 GI:27451838
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2692)
Asada, M., Irie, K., Morimoto, K., Yamada, A., Ikeda, W., Takeuchi, M.
and Takai, Y.
ADIP, a Novel Afadin- and alpha-Actinin-Binding Protein Localized
at Cell-Cell Adherens Junctions
J. Biol. Chem. 278 (6), 4103-4111 (2003)
12446711
2 (bases 1 to 2692)
Asada, M., Irie, K. and Takai, Y.
Submitted (26-JUL-2002) Department of Molecular Biology and
Biochemistry, Osaka University Graduate School of Medicine/Faculty
of Medicine, Yamada-oka 2-2, Suita 565-0871, Japan
Location/Qualifiers
1. .2692
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
80..1927
/codon_start=1
/product="afadin- and alpha-actinin-binding protein ADIP"
/protein_id="AA015015.1"
/db_xref="GI:27451839"
/translation="MGDMVTVPVLCTENKNLSQYTSKTMSPSSLYSQVLCSSVP
LSKNVGVCFVCTGENIQSISYLDQELTFGFPISLYESKSKAEKREINIVAVLNC
MNELLVQRNLLAQESVFQNKJGSDMDHLQSCYAKLKEQETSRRRMIGLQERDR
QLOCKNRSLHQLLNKNEKDEVQKLQNIASRATQYNHDKRKREYNKLERLLHOLVMN

KKDKNIAMDVLNVYVRADGKSGSWRTDKTEARNEDENVKILLINDYEVROKQILMENAE
LKKVLQOMKXEMISLLSPQKKPRERAEDGTGTVAISDIEDDSELSRDSVWGLSCDT
VREQLTNSIRKQWILKSHVEKDNQASKVHSEGLNEEDVTSRODHEQETEKULELEIE
RKEMIAXQOQLLOQJLATTCDDDTSLLRDCYLLBEKERLKEEMTLFKEQKNFERE
RRSFTEAARILGLERKAFEEERASWVKQOFLNMTNFDHNSENVKFLSFASGSGSDPON
LIVHSRPROKKLHVSANGVPACTSKLTKSLPASPSDFRQTHSCVSEHSISVLNIT
PEESKPSEVARESTDQKWSQVSQSRSSREGYSGGSSAFRSHAGDRDLP"

ORIGIN

Query Match		100.0%;	Score 2692;	DB 6;	Length 2692;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 2692;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CGTAGGAGAGTACAGGAGCTGTTGTAAGCGTCCAGCAGCTAGCGCGCTCCTCAGGTAT	60		
Db	1	CGTAGGAGAGTACAGGAGCTGTTGTAAGCGTCCAGCAGCTAGCGCGCTCCTCAGGTAT	60		
Qy	61	CCTGGCTCTGGAACTTGTCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCCTGTGTAC	120		
Db	61	CCTGGCTCTGGAACTTGTCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCCTGTGTAC	120		
Qy	121	AGAAAAAATAATCTCTCAATATACCTCAGAAAAAAGATGTCTCGTCCAGTTTGTATA	180		
Db	121	AGAAAAAATAATCTCTCAATATACCTCAGAAAAAAGATGTCTCGTCCAGTTTGTATA	180		
Qy	181	CTCCAGCAAGTTCGTGCTCTTCACTACCTTATCCAAAAACCTGATGCTGTTTCGG	240		
Db	181	CTCCAGCAAGTTCGTGCTCTTCACTACCTTATCCAAAAACCTGATGCTGTTTCGG	240		
Qy	241	TGCTCTTCGACAGGAGAAACATTTGAACAAGTATTTCTTATCTTGATCAGGAGCTGAC	300		
Db	241	TGCTCTTCGACAGGAGAAACATTTGAACAAGTATTTCTTATCTTGATCAGGAGCTGAC	300		
Qy	301	CACCTTCGGGTTTCTTCTTGTATGAAGAATCCAAAAAGTAAAGAGGCAAGAGAGAATT	360		
Db	301	CACCTTCGGGTTTCTTCTTGTATGAAGAATCCAAAAAGTAAAGAGGCAAGAGAGAATT	360		
Qy	361	AAATATAGTCTGCTGTTCTGAACTGTATGAACGAGCTGCTGCTGCTTCAAGCGGAGAACCT	420		
Db	361	AAATATAGTCTGCTGTTCTGAACTGTATGAACGAGCTGCTGCTGCTTCAAGCGGAGAACCT	420		
Qy	421	GCTGCCCGCAGAGCGTGGAGACACAGAACTTGAAGCTGGGCGAGTGCATGGACACCT	480		
Db	421	GCTGCCCGCAGAGCGTGGAGACACAGAACTTGAAGCTGGGCGAGTGCATGGACACCT	480		
Qy	481	GCAGAGCTGCTACGCCAACTTAAAGGAGCAGTTGGAAACGTCCAGCGGGAGATGATCGG	540		
Db	481	GCAGAGCTGCTACGCCAACTTAAAGGAGCAGTTGGAAACGTCCAGCGGGAGATGATCGG	540		
Qy	541	GCTTCAAGAGAGACAGGCGAGCTGCAGTGCAGAACAGAGGATTTGGCATCAGCTCCGTAA	600		
Db	541	GCTTCAAGAGAGACAGGCGAGCTGCAGTGCAGAACAGAGGATTTGGCATCAGCTCCGTAA	600		
Qy	601	GAATGAGAAAGATGAGGTACAAAATTAACAAAATATCATAGCCAGCGGCTACTCAGTA	660		
Db	601	GAATGAGAAAGATGAGGTACAAAATTAACAAAATATCATAGCCAGCGGCTACTCAGTA	660		
Qy	661	TAATCATGATGTGAAGAGGAAAGGCGGTGAATATAAAGCTAAAGAGCGCCCTGCATCA	720		
Db	661	TAATCATGATGTGAAGAGGAAAGGCGGTGAATATAAAGCTAAAGAGCGCCCTGCATCA	720		
Qy	721	GCTCGTTATGAACAAGAGGATTAACACATAGCCATGGATGTTTAAATTAATGTGGGTG	780		
Db	721	GCTCGTTATGAACAAGAGGATTAACACATAGCCATGGATGTTTAAATTAATGTGGGTG	780		
Qy	781	AGCTGATGGCAAAAGAGGCTCATGGAGATTCAGCAAAAACAGAGCCAGGAATGAAGATGA	840		
Db	781	AGCTGATGGCAAAAGAGGCTCATGGAGATTCAGCAAAAACAGAGCCAGGAATGAAGATGA	840		
Qy	841	GATGTAACAAAATTTCTGTTGAATGATTAATGAGTACCGCCAGAGCAGATCCTGATGGAGAA	900		
Db	841	GATGTAACAAAATTTCTGTTGAATGATTAATGAGTACCGCCAGAGCAGATCCTGATGGAGAA	900		

Qy	901	CGCGGAGCTGAAGAAGTCTCTCAGCAGATGAAGAAGGAGATGATCTCTCTCTCTCTCC	960
Db	901	CGCGGAGCTGAAGAAGTCTCTCAGCAGATGAAGAAGGAGATGATCTCTCTCTCTCTCC	960
Qy	961	TCAGAAGAAGAAGCCAGGGAAGAGCAGAGGAGCGGCA CAGGCACTGTGTCTATCTCCGA	1020
Db	961	TCAGAAGAAGAAGCCAGGGAAGAGCAGAGGAGCGGCA CAGGCACTGTGTCTATCTCCGA	1020
Qy	1021	TATAGAAGATGACTCTGGGAACTGAGCAGAGACAGCGTGTGGGCGCTTCTCTGTGACAC	1080
Db	1021	TATAGAAGATGACTCTGGGAACTGAGCAGAGACAGCGTGTGGGCGCTTCTCTGTGACAC	1080
Qy	1081	TGTGAGAGCAGCTGACAAAACAGCATCAGAAAACAGTGCAGAAATTTTGAAGTCAATGT	1140
Db	1081	TGTGAGAGCAGCTGACAAAACAGCATCAGAAAACAGTGCAGAAATTTTGAAGTCAATGT	1140
Qy	1141	AGAAAACTCGATAACCAAGCTTCGAAGGTACCTCAGAGGGCGCTTAAATGAGGAGACGT	1200
Db	1141	AGAAAACTCGATAACCAAGCTTCGAAGGTACCTCAGAGGGCGCTTAAATGAGGAGACGT	1200
Qy	1201	CATCTCAGCAACAGCCATGACNAAGACTGAGNAACCTGAGCTGAGATTTGAGCGGTG	1260
Db	1201	CATCTCAGCAACAGCCATGAGNAAGACTGAGNAACCTGAGCTGAGATTTGAGCGGTG	1260
Qy	1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGCACACAGCTGTA	1320
Db	1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGCACACAGCTGTA	1320
Qy	1321	TGATGACACCCTCACTGTTGCGAGACTGTTACTTCTGCGAAGAAAGGAAACGCTTTAA	1380
Db	1321	TGATGACACCCTCACTGTTGCGAGACTGTTACTTCTGCGAAGAAAGGAAACGCTTTAA	1380
Qy	1381	AGAAGAGTGAACCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGGCGCAAGCTT	1440
Db	1381	AGAAGAGTGAACCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGGCGCAAGCTT	1440
Qy	1441	TACAGAAGCTGCCATTCGATTGGGTTGGAGAGAAAGCGCTTTGAAGAAGAGCGAGCCAG	1500
Db	1441	TACAGAAGCTGCCATTCGATTGGGTTGGAGAGAAAGCGCTTTGAAGAAGAGCGAGCCAG	1500
Qy	1501	CTGGGTAAGCAGCAGCTTTTAAACATGACGAACTTTTGA CCA CAGAACTCAGAAAATGT	1560
Db	1501	CTGGGTAAGCAGCAGCTTTTAAACATGACGAACTTTTGA CCA CAGAACTCAGAAAATGT	1560
Qy	1561	GAACCTTTTCACTGCTCTCTCAGGAGTCTGATCCAGACAACTTATAGTCCACTCACG	1620
Db	1561	GAACCTTTTCACTGCTCTCTCAGGAGTCTGATCCAGACAACTTATAGTCCACTCACG	1620
Qy	1621	GCCACGGCAAAAGAACTACACAGTGTGGCTAATGGGTTGCCAGCTTGCACATCAAAACT	1680
Db	1621	GCCACGGCAAAAGAACTACACAGTGTGGCTAATGGGTTGCCAGCTTGCACATCAAAACT	1680
Qy	1681	GACTAAATCTCTCTGCTCTCACTTCTA CTT CAGACTTTCGCGCAGACACATTCATGTGT	1740
Db	1681	GACTAAATCTCTCTGCTCTCACTTCTA CTT CAGACTTTCGCGCAGACACATTCATGTGT	1740
Qy	1741	GTCTGACACAGTTCATCAGTGTGCTGAATATAA CTTCTGAGAAAGTAAACCAAGTGA	1800
Db	1741	GTCTGACACAGTTCATCAGTGTGCTGAATATAA CTTCTGAGAAAGTAAACCAAGTGA	1800
Qy	1801	GGTTGCAAGAAAGCAAGCAGGATCAGAGTGGAGCGTGCAGTGCAGGCGCCAGCTCGCGGA	1860
Db	1801	GGTTGCAAGAAAGCAAGCAGGATCAGAGTGGAGCGTGCAGTGCAGGCGCCAGCTCGCGGA	1860
Qy	1861	GGGTGCTACAGCGGATGCTCTCGGCTTTCAGAGCGCTCA CCGGGACCGAGATGACTT	1920
Db	1861	GGGTGCTACAGCGGATGCTCTCGGCTTTCAGAGCGCTCA CCGGGACCGAGATGACTT	1920
Qy	1921	ACCTTAAATCTCGGGCTGAGTCTCTCCAGATGTCGCTAGTACAGAGTGAACACAGG	1980
Db	1921	ACCTTAAATCTCGGGCTGAGTCTCTCCAGATGTCGCTAGTACAGAGTGAACACAGG	1980
Qy	1981	GTGTAGCATAAAGTCACTGCTCTAACTTAAGATGCTCAGAGTGTGTTTGTGGACTTCGC	2040

|||||
1981 GTGTAGCATAAAGTCAGTCGCTTAACCTTAAGATGCTCAGAGTTGTTGTTGGACTTCGC 2040
QY TGTCTTCCCCAAGAGCTGAAATGCTAAAGCTACTTAAAGAGTCAAAAGCTTTGGTTGT 2100
Db TGTCTTCCCCAAGAGCTGAAATGCTAAAGCTACTTAAAGAGTCAAAAGCTTTGGTTGT 2100
QY GTGTTAGTAACAGAGCCCTCGCTCTGTGACTGCGAGGAATGATGCGCTTTGGATGGAA 2160
Db GTGTTAGTAACAGAGCCCTCGCTCTGTGACTGCGAGGAATGATGCGCTTTGGATGGAA 2160
QY ACAGAGCGCTGGAATGATGCTCGCCAGGTACGAGAGAGAGCACTTTTAGGGACTGCT 2220
Db ACAGAGCGCTGGAATGATGCTCGCCAGGTACGAGAGAGAGCACTTTTAGGGACTGCT 2220
QY TCCTGTAACATTAATATTCGTCCTCAAGTGTGGTTCGCAATGGAAGTGTAGCCTTTACT 2280
Db TCCTGTAACATTAATATTCGTCCTCAAGTGTGGTTCGCAATGGAAGTGTAGCCTTTACT 2280
QY TGAATGTATCTAGTATTTTAAACAGAGAGGTTCTATATTTATTTATGTTAGTGTGAT 2340
Db TGAATGTATCTAGTATTTTAAACAGAGAGGTTCTATATTTATTTATGTTAGTGTGAT 2340
QY TTGGGATTACCTCTTTTCATATGTTTGTGCTGTACATAAATATACATGACTATGTTAA 2400
Db TTGGGATTACCTCTTTTCATATGTTTGTGCTGTACATAAATATACATGACTATGTTAA 2400
QY GAGGCTTTAAGGTTTAAAACTTCACACCATCTTTGAGTATAGCATTTTCATGCCAATTAA 2460
Db GAGGCTTTAAGGTTTAAAACTTCACACCATCTTTGAGTATAGCATTTTCATGCCAATTAA 2460
QY AATGTTTTCAGTGGCATGGTGTACAGAGTTAGGACCACTGCCACATGACAGTTAAGA 2520
Db AATGTTTTCAGTGGCATGGTGTACAGAGTTAGGACCACTGCCACATGACAGTTAAGA 2520
QY CTTTATTTTAAAGCCATCTGGSCAATAAAATTTCAAGCCCTTCATAGCTGAGTTCCAG 2580
Db CTTTATTTTAAAGCCATCTGGSCAATAAAATTTCAAGCCCTTCATAGCTGAGTTCCAG 2580
QY ATAAGTAGAAGTAAAGCTTACATTTTGTGAGATTTTAAAGCAATGTTATTTTATTTAT 2640
Db ATAAGTAGAAGTAAAGCTTACATTTTGTGAGATTTTAAAGCAATGTTATTTTATTTAT 2640
QY ATATGTGAATGTTATTAATTTCTAAGAGAAATTTGATTTAGGTAATGAGTAATGGGG 2692
Db ATATGTGAATGTTATTAATTTCTAAGAGAAATTTGATTTAGGTAATGAGTAATGGGG 2692

RESULT 2
AKI73064
LOCUS AKI73064 3424 bp mRNA linear ROD 28-JUL-2004
DEFINITION Mus musculus mRNA for mKIAA0923 protein.
ACCESSION AKI73064
VERSION AKI73064.1 GI:50510712
KEYWORDS FLI_CDNA.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Okazaki,N., Kikuno,R.F., Ohara,R., Inamoto,S., Koseki,H.,
Hiraoka,S., Saga,Y., Seino,S., Nishimura,M., Kaisho,T., Hoshino,K.,
Kitamura,H., Nagase,T., Ohara,O. and Koga,H.
Prediction of the Coding Sequences of Mouse Homologues of KIAA
Gene: IV. The Complete Nucleotide Sequences of 500 Mouse
KIAA-Homologous cDNAs Identified by Screening of Terminal Sequences
of cDNA Clones Randomly Sampled from Size-Fractionated Libraries
DNA Res. 11, 205-218 (2004)
REFERENCE 2 (bases 1 to 3424)
AUTHORS Okazaki,N., Kikuno,R.F., Nagase,T., Ohara,O. and Koga,H.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-2004) Hisashi Koga, Kazusa DNA Research

Institute, Laboratory for Genome Informatics; 2-6-7
Kazusa-kanatari, Kisarazu, Chiba 292-0818, Japan
(E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT The CREATE program supported by Japan science and technology
corporation; cDNA full insert sequencing; Kazusa DNA Research
Institute; cDNA library construction, clone selection and 5'- &
3'-end one pass sequencing.
FEATURES
Location/Qualifiers
1..3424
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="mph02954"
/tissue_type="embryonic tail"
/note="Vector:modified pBC SK+"
1..3424
/gene="mKIAA0923"
/db_xref="GeneID:236"
/gene="mKIAA0923"
/inference="non-experimental evidence, no additional
details recorded"
/note="CDS is predicted by in silico analysis. Start codon
is not identified."
/codon_start=1
/product="mKIAA0923 protein"
/protein_id="BAD32342.1"
/db_xref="GI:50510713"
/translation="AASSGILALELAMGDMVTDPVLTENKNLSOYVTSKMPSS
LYSOVLSSVPLSKNVHGVFGVCTGENIEQSISYLDDELTPGFPSELSKSKA
KRELNIVAVLNCMELLVLQRNLLAQESVETQNLKSGDMDHLOSVCYAKLKEQETS
RREMIGQERDRQLOKRNLSUHLNKEDEKQVQLQNIASRTAQTHDHYKKREYN
KLKRLHQLNKKDKNIADVLNVGRADRGSRWRTDKTEARNEEDMYKILLNDEY
YRQGIILMENAELKKVLQOMKEMISLLSPQKKPRERAEDGTGTAISDIEDDSGEL
SRDSVWGLSCDVTREQLTNSIRKQWIRILKSHVEKLDNOASKVHSEGLNEEDVISRQDH
BOETKLELETERCKEMIKAQQLQOOLATTCDDDTTSLARDCYLLBEKRLKEWT
LFEQKNFERERSFTAAIRLGLERKAFEEBASVWQFLLNMTNFDHQNSENVKL
FSFSGSDPNLIVHSRPROKHLHSVANGVPACTSKTLSPASPTSDFRQTHSCV
SEHSISVLNITPEESKPESEVARESTDQKWSVQSRPSSREGCYSCSAPRSHAGDRD
DLP"

ORIGIN
Query Match 100.0%; Score 2692; DB 6; Length 3424;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2692; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGTAGGAGAGTGACAGAGAGCTGTTGTAAGCGTCGCAGCACCTGAGCGGCTCTCTCAGGTAT 60
Db |||||
QY 193 CGTAGGAGAGTGACAGAGAGCTGTTGTAAGCGTCGCAGCACCTGAGCGGCTCTCTCAGGTAT 252
Db |||||
QY 61 CCTGGCTCTGGAACCTTGTATGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTAC 120
Db |||||
QY 253 CCTGGCTCTGGAACCTTGTATGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTAC 312
Db |||||
QY 121 AGAAAAAATAATCTCTCAATATACCTCAGAACAAAGATGTCTCGTCCAGTTTGTGA 180
Db |||||
QY 313 AGAAAAAATAATCTCTCAATATACCTCAGAACAAAGATGTCTCGTCCAGTTTGTGA 372
Db |||||
QY 181 CTCCAGCAAGTCTGTGCTCTTCTCAGTACCTTTATCCAAAAACGTGCATGTTGTTTCGG 240
Db |||||
QY 373 CTCCAGCAAGTCTGTGCTCTTCTCAGTACCTTTATCCAAAAACGTGCATGTTGTTTCGG 432
Db |||||
QY 241 TGTCTTCTGCACAGAGAGAACATTGAAACAAGTATTTCTCTATCTTGATCAGGAGCTGAC 300
Db |||||
QY 433 TGTCTTCTGCACAGAGAGAACATTGAAACAAGTATTTCTCTATCTTGATCAGGAGCTGAC 492
Db |||||
QY 301 CACCTTCGGGTTTCTCTCTTGTATGAGATCCAAAAAGTAAAGAGGCAAGAGAGATT 360
Db |||||
QY 493 CACCTTCGGGTTTCTCTCTTGTATGAGATCCAAAAAGTAAAGAGGCAAGAGAGATT 552
Db |||||
QY 361 AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTGCTGCTTTCAGCGGAAGAACCT 420
Db |||||
QY 553 AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTGCTGCTTTCAGCGGAAGAACCT 612
Db |||||

Qy	421	GCTGCCACGAGAGCGTGGAGACACAGAACTTGAAGCTGGCGAGTGACATGGACCACCT	480
Db	613	GCTGCCACGAGAGCGTGGAGACACAGAACTTGAAGCTGGCGAGTGACATGGACCACCT	672
Qy	481	GCAGAGCTGCTACGCGCAAACTTAAAGGACAGTGTGAAACGCTCCAGCGGGGAGATGATCGG	540
Db	673	GCAGAGCTGCTACGCGCAAACTTAAAGGACAGTGTGAAACGCTCCAGCGGGGAGATGATCGG	732
Qy	541	GCTTCAAGAGAGACAGCGAGCTGCAGTCAAGAACAGGAGTTTGCATCAGCTCCTGAA	600
Db	733	GCTTCAAGAGAGACAGCGAGCTGCAGTCAAGAACAGGAGTTTGCATCAGCTCCTGAA	792
Qy	601	GAATGAGAAAGATGAGGTACAAAATTAACAAAATATCATAGCCAGCGGGCTACTCAGTA	660
Db	793	GAATGAGAAAGATGAGGTACAAAATTAACAAAATATCATAGCCAGCGGGCTACTCAGTA	852
Qy	661	TAATCATGATGTGAAGAGGAAGGCGTGAATATAATAAGCTAAAGGAGCGCCTGCATCA	720
Db	853	TAATCATGATGTGAAGAGGAAGGCGTGAATATAATAAGCTAAAGGAGCGCCTGCATCA	912
Qy	721	GCTGCTTATGAACAAGAGGATATAAACATAGCCATGAGTGTTTAAATATGTGGGTG	780
Db	913	GCTGCTTATGAACAAGAGGATATAAACATAGCCATGAGTGTTTAAATATGTGGGTG	972
Qy	781	AGCTGATGCGCAAAACGAGGCTCATGGAGACTGACAAAACAGAACGACAGGAATGAAGTGA	840
Db	973	AGCTGATGCGCAAAACGAGGCTCATGGAGACTGACAAAACAGAACGACAGGAATGAAGTGA	1032
Qy	841	GATGTACAAAATCTGTTGAATGATTAATGAGTACCGCCAGAACAGATCTCTGATGGAGAA	900
Db	1033	GATGTACAAAATCTGTTGAATGATTAATGAGTACCGCCAGAACAGATCTCTGATGGAGAA	1092
Qy	901	CGCGAGCTGAAGAGTCTCCAGCAGATGAAGAACAGAGATGATCTCTCTCTCTCC	960
Db	1093	CGCGAGCTGAAGAGTCTCCAGCAGATGAAGAACAGAGATGATCTCTCTCTCTCC	1152
Qy	961	TCAGAAAGAAAGCCACGGGAAAGACAGAGAGCGGACAGCGCACTGTGCTATCTCCGA	1020
Db	1153	TCAGAAAGAAAGCCACGGGAAAGACAGAGAGCGGACAGCGCACTGTGCTATCTCCGA	1212
Qy	1021	TATAGAAGATGACTCTGCGGAACTGAGCAGAGACAGCGGTGTGGGCTTTCTCTGTGACAC	1080
Db	1213	TATAGAAGATGACTCTGCGGAACTGAGCAGAGACAGCGGTGTGGGCTTTCTCTGTGACAC	1272
Qy	1081	TGTGAGAGCAGCTGACAAACACATCAGAAACAGTGGAGAAATTTGAAAAGTCACTGT	1140
Db	1273	TGTGAGAGCAGCTGACAAACACATCAGAAACAGTGGAGAAATTTGAAAAGTCACTGT	1332
Qy	1141	AGAAAACTCGATAAACCAAGCTTCGAAGGTACACTCAGAGGGCTTAAATGAGGAGGACGT	1200
Db	1333	AGAAAACTCGATAAACCAAGCTTCGAAGGTACACTCAGAGGGCTTAAATGAGGAGGACGT	1392
Qy	1201	CATCTCAGCAACAGACCATGAGCAAGAGACTGAGAAACTGGAGACTGGAGATGAGCGGTG	1260
Db	1393	CATCTCAGCAACAGACCATGAGCAAGAGACTGAGAAACTGGAGACTGGAGATGAGCGGTG	1452
Qy	1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGGCCACCACTGTA	1320
Db	1453	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGGCCACCACTGTA	1512
Qy	1321	TGATGACACCACTCCTCTGTTGCGAGACTGTACTTGTCTGGAAGAAAAGGAAACCCCTTAA	1380
Db	1513	TGATGACACCACTCCTCTGTTGCGAGACTGTACTTGTCTGGAAGAAAAGGAAACCCCTTAA	1572
Qy	1381	AGAAAGTGGACCCCTTTTAAAGAGCAAAAAAGAAATTTTGAAGAGAAAGGCGGAAGCTT	1440
Db	1573	AGAAAGTGGACCCCTTTTAAAGAGCAAAAAAGAAATTTTGAAGAGAAAGGCGGAAGCTT	1632
Qy	1441	TACAGAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGCTTTCAGAGAGAGCGGCGAG	1500
Db	1633	TACAGAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGCTTTCAGAGAGAGCGGCGAG	1692
Qy	1501	CTGGGTAAAGCAGCAGTGTGTAAACATGACGAACTTTGACCAACCAAGAACTCAGAAAAATGT	1560

Db	1693	CTGGGTAAAGCAGCAGTGTGTAAACATGACGAACTTTGACCAACCAAGAACTCAGAAAAATGT	1752
Qy	1561	GAAACTTTTTCAGTGCCCTTCTCAGGAAGTTCTGATCCAGACAAATCTTATAGTCCACTCAGC	1620
Db	1753	GAAACTTTTTCAGTGCCCTTCTCAGGAAGTTCTGATCCAGACAAATCTTATAGTCCACTCAGC	1812
Qy	1621	GCCACGCGAAAGAGCTACACAGTGTGGCTTAATGGGGTGCCAGCTTGCAATCAAAACT	1680
Db	1813	GCCACGCGAAAGAGCTACACAGTGTGGCTTAATGGGGTGCCAGCTTGCAATCAAAACT	1872
Qy	1681	GACTAAATCTCTCTCTCGCTCAGCTTCTACTTTCAGACTTTTCGCCAGACACATTCATGTGT	1740
Db	1873	GACTAAATCTCTCTCTCGCTCAGCTTCTACTTTCAGACTTTTCGCCAGACACATTCATGTGT	1932
Qy	1741	GTCTGAAACACAGTTCATCAGTGTGCTGAATATAAATCTCTGAAAGAAAGTAAACCAAGTGA	1800
Db	1933	GTCTGAAACACAGTTCATCAGTGTGCTGAATATAAATCTCTGAAAGAAAGTAAACCAAGTGA	1992
Qy	1801	GGTTGCAAGAGAAAGCACCGATCAGAGTGGAGCGTGCAGTTCGAGGCCAGCTCGCCGGA	1860
Db	1993	GGTTGCAAGAGAAAGCACCGATCAGAGTGGAGCGTGCAGTTCGAGGCCAGCTCGCCGGA	2052
Qy	1861	GGGTGTCTACAGCGGATGCTCTCTCGGCTTTCAGGAGCGCTCACGGGACCGAGATGACTT	1920
Db	2053	GGGTGTCTACAGCGGATGCTCTCTCGGCTTTCAGGAGCGCTCACGGGACCGAGATGACTT	2112
Qy	1921	ACCTTAAATGTGCGGCTGCAGTGTCTTCCAGATGTGCGCTAGAGGAGTTGACACAGG	1980
Db	2113	ACCTTAAATGTGCGGCTGCAGTGTCTTCCAGATGTGCGCTAGAGGAGTTGACACAGG	2172
Qy	1981	GTGTAGCATAAAGTCAAGTGTCTTAAAGTGTCTCAGAGTGTGTTTGTGGACTTTCGC	2040
Db	2173	GTGTAGCATAAAGTCAAGTGTCTTAAAGTGTCTCAGAGTGTGTTTGTGGACTTTCGC	2232
Qy	2041	TGCTTTTCCCAAGAGCTGAAATGTACTTAAAGAGTGTGAAAGTGTGTTTGTGGTGT	2100
Db	2233	TGCTTTTCCCAAGAGCTGAAATGTACTTAAAGAGTGTGAAAGTGTGTTTGTGGTGT	2292
Qy	2101	GTGTAGTAAACAGAGCCCTGCTGTGACTGTGAGTGTGAGGAAATGATGGCGTTTGGATGAA	2160
Db	2293	GTGTAGTAAACAGAGCCCTGCTGTGACTGTGAGTGTGAGGAAATGATGGCGTTTGGATGAA	2352
Qy	2161	ACAGAGCGCTGGAATGATTGCTCGCAGGTACCGAGAAAGAGCACCTTTTAGGAGCTGGT	2220
Db	2353	ACAGAGCGCTGGAATGATTGCTCGCAGGTACCGAGAAAGAGCACCTTTTAGGAGCTGGT	2412
Qy	2221	TCCTGTAAACATTTAAATATTTCTGTTCCCAAGTGTGGTGTGCAATTTGAAAGTGTAGCTTTACT	2280
Db	2413	TCCTGTAAACATTTAAATATTTCTGTTCCCAAGTGTGGTGTGCAATTTGAAAGTGTAGCTTTACT	2472
Qy	2281	TGAATGTATCTGTAGATTTTAAACAGCAGGTCTTATATTTATTTATGTTTAGTGTGAT	2340
Db	2473	TGAATGTATCTGTAGATTTTAAACAGCAGGTCTTATATTTATTTATGTTTAGTGTGAT	2532
Qy	2341	TTTGGGATTAACCTTTTATATGTTTGTCTGTACATAAAATACATGACTATGTGTTAA	2400
Db	2533	TTTGGGATTAACCTTTTATATGTTTGTCTGTACATAAAATACATGACTATGTGTTAA	2592
Qy	2401	GAGGCTTTAAAGTGTAAACCTTACACCATGCTTGGATATAGCATTTTCATGCAATTA	2460
Db	2593	GAGGCTTTAAAGTGTAAACCTTACACCATGCTTGGATATAGCATTTTCATGCAATTA	2652
Qy	2461	AATGTTTTTCAGTGGCATGGTGTTCAGAGGTTAGGACCACTGCCACATCAGCAGTTAAGA	2520
Db	2653	AATGTTTTTCAGTGGCATGGTGTTCAGAGGTTAGGACCACTGCCACATCAGCAGTTAAGA	2712
Qy	2521	CTTTATTTTTAAGCCATCTGGGCAATAAAATTTCAAAGCCCCCTTCATAAGCTGAGTTTCA	2580
Db	2713	CTTTATTTTTAAGCCATCTGGGCAATAAAATTTCAAAGCCCCCTTCATAAGCTGAGTTTCA	2772
Qy	2581	ATAACTAGAACTACTAACCGTTTACATTTTGAAGATTTTAAAGCATTTGATTTTATTTAT	2640

Qy 541 GCTTCAAGAGAGACAGGCGAGCTGCAGTGCAGAAACAGAGAGTTTGCATCAGCTCCTGAA 600
Db |||||
Qy 701 GCITTCAGAGAGAGACAGGCGAGCTGCAGTGCAGAAACAGAGAGTTTGCATCAGCTCCTGAA 760
Db |||||
Qy 601 GAATGAGAAAGATCAGGTACAAAATTTACAAAATATCATAGCCAGCGGGCTACTCAGTA 660
Db |||||
Qy 761 GAATGAGAAAGATCAGGTACAAAATTTACAAAATATCATAGCCAGCGGGCTACTCAGTA 820
Db |||||
Qy 661 TAATCATGATGTGAAGAGAAAGGAGCGTGAATATAATAAGCTAAAGGAGCGCTGCATCA 720
Db |||||
Qy 821 TAATCATGATGTGAAGAGAAAGGAGCGTGAATATAATAAGCTAAAGGAGCGCTGCATCA 880
Db |||||
Qy 721 GCTCGTTATGAACAGAGAGATTAATAACATAGCCATCGATGTTTTAAATTTATGCGGTGC 780
Db |||||
Qy 881 GCTCGTTATGAACAGAGAGATTAATAACATAGCCATCGATGTTTTAAATTTATGCGGTGC 940
Db |||||
Qy 781 AGCTGATGGCAAAACGAGGCTCATGGAGACTGCACAAAACAGAAAGCCAGGAATGAAGATGA 840
Db |||||
Qy 941 AGCTGATGGCAAAACGAGGCTCATGGAGACTGCACAAAACAGAAAGCCAGGAATGAAGATGA 1000
Db |||||
Qy 841 GATGTACAAAATTTCTGTTGAATGATTTATGAGTACCGCCAGAGAGATCCTGATGGAGAA 900
Db |||||
Qy 1001 GATGTACAAAATTTCTGTTGAATGATTTATGAGTACCGCCAGAGAGATCCTGATGGAGAA 1060
Db |||||
Qy 901 CGCGAGCTGAAGAGGTCCTCCAGCAGATGAAGAAGAGAGATGATCTCTCCTGTCTCC 960
Db |||||
Qy 1061 CGCGAGCTGAAGAGGTCCTCCAGCAGATGAAGAAGAGAGATGATCTCTCCTGTCTCC 1120
Db |||||
Qy 961 TCAGAAGAAGAGCCAGGGAAAGAGCAGAGAGCGGCACAGGCACTGTGCTATCTCCGA 1020
Db |||||
Qy 1121 TCAGAAGAAGAGCCAGGGAAAGAGCAGAGAGCGGCACAGGCACTGTGCTATCTCCGA 1180
Db |||||
Qy 1021 TATAGAGATGATCTCTGGGAACTAGCAGAGAGACAGCGTGTGGGCGCTTTCCTGTGACAC 1080
Db |||||
Qy 1181 TATAGAGATGATCTCTGGGAACTAGCAGAGAGACAGCGTGTGGGCGCTTTCCTGTGACAC 1240
Db |||||
Qy 1081 TGTGAGAGAGAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAGAGTCATGT 1140
Db |||||
Qy 1241 TGTGAGAGAGAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAGAGTCATGT 1300
Db |||||
Qy 1141 AGAAAACTCGATAACCAAGCTTCGAAGGTTACACTCAGAGGGCGCTTAATGAGGAGGACGT 1200
Db |||||
Qy 1301 AGAAAACTCGATAACCAAGCTTCGAAGGTTACACTCAGAGGGCGCTTAATGAGGAGGACGT 1360
Db |||||
Qy 1201 CATCTCAGCAAGACCAATGAGCAAGAGATGAGAGAACTGGAGCTGGAGATTTGAGCGGTG 1260
Db |||||
Qy 1361 CATCTCAGCAAGACCAATGAGCAAGAGATGAGAGAACTGGAGCTGGAGATTTGAGCGGTG 1420
Db |||||
Qy 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACTGTGA 1320
Db |||||
Qy 1421 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACTGTGA 1480
Db |||||
Qy 1321 TGATGACACCACTCAGCTGTTGCGAGACTGTTACTTGTGGAAGAAAGGAAAGCCCTTAA 1380
Db |||||
Qy 1481 TGATGACACCACTCAGCTGTTGCGAGACTGTTACTTGTGGAAGAAAGGAAAGCCCTTAA 1540
Db |||||
Qy 1381 AGAAGAGTGAACCTTTTAAAGAGCAAAAAGAAATTTTGAAGAGAAAGGCGCAAGCTT 1440
Db |||||
Qy 1541 AGAAGAGTGAACCTTTTAAAGAGCAAAAAGAAATTTTGAAGAGAAAGGCGCAAGCTT 1600
Db |||||
Qy 1441 TACAGAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGTTTGAAGAAGAGCGAGCCAG 1500
Db |||||
Qy 1601 TACAGAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGTTTGAAGAAGAGCGAGCCAG 1660
Db |||||
Qy 1501 CTGGGTAAAGCAGCAGTTTTTAAACATGACGAACTTTTGACCAACAGAACTCAGAAAAATGT 1560
Db |||||
Qy 1661 CTGGGTAAAGCAGCAGTTTTTAAACATGACGAACTTTTGACCAACAGAACTCAGAAAAATGT 1720
Db |||||
Qy 1561 GAAACTTTTTCAGTGCCTCTCAGGAGTTCTGATCCAGAGAAATCTTATAGTCCACTCAGC 1620
Db |||||
Qy 1721 GAAACTTTTTCAGTGCCTCTCAGGAGTTCTGATCCAGAGAAATCTTATAGTCCACTCAGC 1780
Db |||||
Qy 1621 GCCACGGCAAAAGAGCTACACAGTGTGGCTAAATGGGGTGCAGCTTGCACATCAAAACT 1680

Db 1781 GCCACGGCAAAAGAGCTACACAGTGTGGCTAATGGGGTCCAGCTTGCACATCAAAACT 1840
Qy |||||
Qy 1681 GACTAAATCTCTTCTCGCTCCACCTTCTACTTCAGACTTTTCGCCAGACACATTCATGTGT 1740
Db |||||
Qy 1841 GACTAAATCTCTTCTCGCTCCACCTTCTACTTCAGACTTTTCGCCAGACACATTCATGTGT 1900
Db |||||
Qy 1741 GTCTGAACACAGTTCCATCAGTGTGCTGAATATAAATCCCTGAAGAAAGTAAACCAAGTGA 1800
Db |||||
Qy 1901 GTCTGAACACAGTTCCATCAGTGTGCTGAATATAAATCCCTGAAGAAAGTAAACCAAGTGA 1960
Db |||||
Qy 1801 GGTTCGAAGAGAAAGACCGGATCAGAAAGTGAGCGTGCAGTGCAGGGCCAGCTCGCGGA 1860
Db |||||
Qy 1961 GGTTCGAAGAGAAAGACCGGATCAGAAAGTGAGCGTGCAGTGCAGGGCCAGCTCGCGGA 2020
Db |||||
Qy 1861 GGGTGTCTACAGCGGATGCTCTCGGCTTCAGAGCGCTCACGGGGACCGAGATGACTT 1920
Db |||||
Qy 2021 GGGTGTCTACAGCGGATGCTCTCGGCTTCAGAGCGCTCACGGGGACCGAGATGACTT 2080
Db |||||
Qy 1921 ACCTTAAATGTGCGGGCTGCAGTGTCTCCAGATGTGCGCTAGAGAGTTGACACAGG 1980
Db |||||
Qy 2081 ACCTTAAATGTGCGGGCTGCAGTGTCTCCAGATGTGCGCTAGAGAGTTGACACAGG 2140
Db |||||
Qy 1981 GTGTAGCATAAAGTTCAGTGTCTTAAAGTTCAGAGTTGTGTTTGTGGACTTCGC 2040
Db |||||
Qy 2141 GTGTAGCATAAAGTTCAGTGTCTTAAAGTTCAGAGTTGTGTTTGTGGACTTCGC 2200
Db |||||
Qy 2041 TGTCTTCCCCCAAGAGCTGAAATGTAGCTACTTAAAGGATGCAAGCTTTGGTTGT 2100
Db |||||
Qy 2201 TGTCTTCCCCCAAGAGCTGAAATGTAGCTACTTAAAGGATGCAAGCTTTGGTTGT 2260
Db |||||
Qy 2101 GTGTAGTAAACAGAGCCCTGGCTCTGTGACTGCAGGAATGATGCGCTTGGATGCAA 2160
Db |||||
Qy 2261 GTGTAGTAAACAGAGCCCTGGCTCTGTGACTGCAGGAATGATGCGCTTGGATGCAA 2320
Db |||||
Qy 2161 ACAGAGCGCTGGAATGATTCCTCGCCAGGTACCGAGAGAGACACTTTTAGGACTTGGT 2220
Db |||||
Qy 2321 ACAGAGCGCTGGAATGATTCCTCGCCAGGTACCGAGAGAGACACTTTTAGGACTTGGT 2380
Db |||||
Qy 2221 TCCTGTAAACATTAATATTCGTCCTCCAGTGTGGTATGGCAATGGAGTGTAGCCTTACT 2280
Db |||||
Qy 2381 TCCTGTAAACATTAATATTCGTCCTCCAGTGTGGTATGGCAATGGAGTGTAGCCTTACT 2440
Db |||||
Qy 2281 TGAATGTATACCTGTAGATTTTAAACAGCAGGTTCTATATTTATTTAGTTAGTGTGAT 2340
Db |||||
Qy 2441 TGAATGTATACCTGTAGATTTTAAACAGCAGGTTCTATATTTATTTAGTTAGTGTGAT 2500
Db |||||
Qy 2341 TTTGGGATTAACCTCTTTTCATATGTTTGTCTGTACATAAAATATACATGACTATGTAA 2400
Db |||||
Qy 2501 TTTGGGATTAACCTCTTTTCATATGTTTGTCTGTACATAAAATATACATGACTATGTAA 2560
Db |||||
Qy 2401 GAGGCTTTAAGGTTTAAAACTTCACACCATGCTTGTAGTATAGCATTTTCATGCCAAATTA 2460
Db |||||
Qy 2561 GAGGCTTTAAGGTTTAAAACTTCACACCATGCTTGTAGTATAGCATTTTCATGCCAAATTA 2620
Db |||||
Qy 2461 AATGTTTTTCAGTGCAGTGTGTTTACAGAGGTTTAGGACCACTGCACATGACAGTTAAGA 2520
Db |||||
Qy 2621 AATGTTTTTCAGTGCAGTGTGTTTACAGAGGTTTAGGACCACTGCACATGACAGTTAAGA 2680
Db |||||
Qy 2521 CTTTATTTTTAAGCCATCTGGGCAATAAAAAATTTCAAAGCCCCCTTCATAAGCTGAGTTGAG 2580
Db |||||
Qy 2681 CTTTATTTTTAAGCCATCTGGGCAATAAAAAATTTCAAAGCCCCCTTCATAAGCTGAGTTGAG 2740
Db |||||
Qy 2581 ATAACTAGAACTACTAAAGTTTACATTTTTCAGATTTTAAAGCATTTGATTTTATTTAT 2640
Db |||||
Qy 2741 ATAACTAGAACTACTAAAGTTTACATTTTTCAGATTTTAAAGCATTTGATTTTATTTAT 2800
Db |||||
Qy 2641 ATATGCAATGTTATATTTCTAAAGAGGAATATTTGATTTATGGAGTAATGGGG 2692
Db |||||
Qy 2801 ATATGCAATGTTATATTTCTAAAGAGGAATATTTGATTTATGGAGTAATGGGG 2852
Db |||||

BC031527	BC031527	3410 bp	mRNA	linear	ROD 29-JUN-2004
LOCUS	Mus musculus synovial sarcoma, X breakpoint 2 interacting protein,				
DEFINITION	mRNA (CDNA clone MGC:28268 IMAGE:4008624), complete cds.				
ACCESSION	BC031527				
VERSION	BC031527.1	GI:21594536			
KEYWORDS	MGC.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 3410)				
	Strausberg, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Klausner, R.D., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udwin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullighy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywicki, M.I., Skalek, U., Smalhus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.				
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)				
PUBMED	12477932				
REFERENCE	2 (bases 1 to 3410)				
AUTHORS	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk Email: c9apbs-remail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (L1NL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: amg@bcm.tmc.edu Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.				
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/L1NL at: http://image.llnl.gov Series: IRAK Plate: 36 Row: k Column: 10 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 20270272. Location/Qualifiers 1..3410 /organism="Mus musculus" /mol_type="mRNA" /strain="Czech II" /db_xref="taxon:10090" /clone="MGC:28268 IMAGE:4008624" /tissue type="Mammary tumor metastasized to lung. MMTV-LTR/Mnt1 model. Expression driven by an MMTV-LTR enhancer." /clone_lib="NCI_CGAP_Lu30"				

gene	/lab_host="DH10B" /note="Vector: pCMV-SPORT6" 1..3410 /gene="synonym: Adip" /db_xref="GeneID:99167" /db_xref="MGI:2139150" 248..2095 /gene="Ssx2ip" /codon_start=1 /product="synovial sarcoma, X breakpoint 2 interacting protein" protein_id="AAH31527.1" /db_xref="GI:21594537" /db_xref="GeneID:99167" /db_xref="MGI:2139150"	CDS	1 248 2095
QY	1	CGTAGGAGATGACAGGAGCTGTTGTAAGCGTCGACGACCTGAGCGGCTCTCTCAGGATAT	60
DB	169	CGTAGGAGATGACAGGAGCTGTTGTAAGCGTCGACGACCTGAGCGGCTCTCTCAGGATAT	228
QY	61	CCTGCTCTGGAACCTTGTATGGAGATTTGATGATCTGTGACAGATTCAGTTCTGTGTAC	120
DB	229	CCTGCTCTGGAACCTTGTATGGAGATTTGATGATCTGTGACAGATTCAGTTCTGTGTAC	288
QY	121	AGAAACAAAATCTCTCAATATACCTCAGAACAAAGATGTTCTCGTCCAGTTGTA	180
DB	289	AGAAACAAAATCTCTCAATATACCTCAGAACAAAGATGTTCTCGTCCAGTTGTA	348
QY	181	CTCCAGCAAGTCTGTGCTCTTCTCAGTACCTTTATCCAAAAACGTCATGGTTTCGG	240
DB	349	CTCCAGCAAGTCTGTGCTCTTCTCAGTACCTTTATCCAAAAACGTCATGGTTTCGG	408
QY	241	TGCTTTCTGCACAGGAGAGAACATTTGAACAAAGTATTTCTCTATCTTATCAGGAGCTGAC	300
DB	409	TGCTTTCTGCACAGGAGAGAACATTTGAACAAAGTATTTCTCTATCTTATCAGGAGCTGAC	468
QY	301	CACCTTCGGGTTTCTTCTTGTATGAAGATCCAAAGTAAAGAGGCAAGAGAGAAAT	360
DB	469	CACCTTCGGGTTTCTTCTTGTATGAAGATCCAAAGTAAAGAGGCAAGAGAGAAAT	528
QY	361	AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTGCTGCTTTCAGCGGAAGACT	420
DB	529	AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTGCTGCTTTCAGCGGAAGACT	588
QY	421	GCTGCCCGAGGAGCGTGGAGACACAGAACTTGAAGCTGGGCGAGTGATGACACCACT	480
DB	589	GCTGCCCGAGGAGCGTGGAGACACAGAACTTGAAGCTGGGCGAGTGATGACACCACT	648
QY	481	GCAGAGCTGTACCCCAACTTAAAGGAGCAGTTGGAACAGTCCAGCGGGAGATGATCGG	540
DB	649	GCAGAGCTGTACCCCAACTTAAAGGAGCAGTTGGAACAGTCCAGCGGGAGATGATCGG	708
QY	541	GCTTCAAGAGAGACAGGAGCTGTCAGTGCACAAACAGAGAGTTTGCATCAGCTCCCTGAA	600
DB	709	GCTTCAAGAGAGACAGGAGCTGTCAGTGCACAAACAGAGAGTTTGCATCAGCTCCCTGAA	768

QY 601 GAATGAGAAAGATGAGGTACAAAATTAACAAAATATCATAGCCAGCGGGCTACTCAGTA 660
DB |||||||
QY 769 GAATGAGAAAGATGAGGTACAAAATTAACAAAATATCATAGCCAGCGGGCTACTCAGTA 828
DB |||||||
QY 661 TAATCATGATGTGAAGAGGAAGGCGGTGAATATAATAAGCTAAAGGAGCGCCTGCATCA 720
DB |||||||
QY 829 TAATCATGATGTGAAGAGGAAGGCGGTGAATATAATAAGCTAAAGGAGCGCCTGCATCA 888
DB |||||||
QY 721 GCTGCTTATGAACAAGAGGATTAACACATAGCCATGGATGTTTTAAATATATGTTGGGTG 780
DB |||||||
QY 889 GCTGCTTATGAACAAGAGGATTAACACATAGCCATGGATGTTTTAAATATATGTTGGGTG 948
DB |||||||
QY 781 AGCTGATGCGCAACGAGGCTCATGAGGACTGACAAAACAGAACGACGAGGAATGAAGATGA 840
DB |||||||
QY 949 AGCTGATGCGCAACGAGGCTCATGAGGACTGACAAAACAGAACGACGAGGAATGAAGATGA 1008
DB |||||||
QY 841 GATGTACAAAATTTCTGTTGAATGATATGAGTACCGCCAGAACGAGATCTCTGATGGAGAA 900
DB |||||||
QY 1009 GATGTACAAAATTTCTGTTGAATGATATGAGTACCGCCAGAACGAGATCTCTGATGGAGAA 1068
DB |||||||
QY 901 CGCGAGCTGAAGAGGTCCTCCAGCAGATGAAGAGGAGATGATCTCTCTCTGTCTCC 960
DB |||||||
QY 1069 CGCGAGCTGAAGAGGTCCTCCAGCAGATGAAGAGGAGATGATCTCTCTCTGTCTCC 1128
DB |||||||
QY 961 TCAGAAGAAGAGCCAGGGAAAGAGCAGAGGACGCGACAGGCACTGTTGCTATCTCCGA 1020
DB |||||||
QY 1129 TCAGAAGAAGAGCCAGGGAAAGAGCAGAGGACGCGACAGGCACTGTTGCTATCTCCGA 1188
DB |||||||
QY 1021 TATAGAAGATGACTCTGGGGAACTGAGCAGAGACAGCGTGTGGGGCCCTTCTCTGTGAC 1080
DB |||||||
QY 1189 TATAGAAGATGACTCTGGGGAACTGAGCAGAGACAGCGTGTGGGGCCCTTCTCTGTGAC 1248
DB |||||||
QY 1081 TGTGAGAGAGAGCTGACAAACAGCATCAGGAACAGTGGAGAAATTTTGAAGATCATGT 1140
DB |||||||
QY 1249 TGTGAGAGAGAGCTGACAAACAGCATCAGGAACAGTGGAGAAATTTTGAAGATCATGT 1308
DB |||||||
QY 1141 AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGGCCCTTAATGAGGAGGAGCT 1200
DB |||||||
QY 1309 AGAAAACTCGATAACCAAGCTTCGAAGGTGCATCTCAGAGGGCCCTTAATGAGGAGGAGCT 1368
DB |||||||
QY 1201 CATCTCAGCACAAGACCAATGAGCAAGAGACTGAGAAACTGGAGCTGGAGATTGAGCGGTG 1260
DB |||||||
QY 1369 CATCTCAGCACAAGACCAATGAGCAAGAGACTGAGAAACTGGAGCTGGAGATTGAGCGGTG 1428
DB |||||||
QY 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGSCCACCACCGTGA 1320
DB |||||||
QY 1429 TAAAGAAATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACCGTGA 1488
DB |||||||
QY 1321 TGATGACACCACTCAGCTGTTGCGAGACTGTTACTTGTGGAAGAAAGGAACGCCTTAA 1380
DB |||||||
QY 1489 TGATGACACCACTCAGCTGTTGCGAGACTGTTACTTGTGGAAGAAAGGAACGCCTTAA 1548
DB |||||||
QY 1381 AGAAGAGTGGACCTTTTAAAGAGCAAAAAAGAAATTTTGAAGAGAAAGCGCAAGCTT 1440
DB |||||||
QY 1549 AGAAGAGTGGACCTTTTAAAGAGCAAAAAAGAAATTTTGAAGAGAAAGCGCAAGCTT 1608
DB |||||||
QY 1441 TACAGAAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGCTTTGAGNAGAGCAGCCAG 1500
DB |||||||
QY 1609 TACAGAAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGCTTTGAGNAGAGCAGCCAG 1668
DB |||||||
QY 1501 CTGGTAAAGCAGCAGTGTGTTTAAACATGACCAACTTTTGAACCAACAGAACTCAGAAAAATGT 1560
DB |||||||
QY 1669 CTGGTAAAGCAGCAGTGTGTTTAAACATGACCAACTTTTGAACCAACAGAACTCAGAAAAATGT 1728
DB |||||||
QY 1561 GAAACTTTTCACTGCTCTCTCAGGAAGTTCTGATCCAGACAATCTTATAGTCCACTCAG 1620
DB |||||||
QY 1729 GAAACTTTTCACTGCTCTCTCAGGAAGTTCTGATCCAGACAATCTTATAGTCCACTCAG 1788
DB |||||||
QY 1621 GCCAGGCAAAAGAGCTTACACAGTGTGGCTTAATGGGTGCGAGCTTGCCATCAAACT 1680
DB |||||||
QY 1789 GCCAGGCAAAAGAGCTTACACAGTGTGGCTTAATGGGTGCGAGCTTGCCATCAAACT 1848
DB |||||||
QY 1681 GACTAAATCTCTCTCTGCTCCTCACCTTCTACTTCAGACTTTCCGCCAGACACATTCATGTGT 1740

DB |||||||
QY 1849 GACTAAATCTCTCTGCTCCTCACCTTCTACTTCAGACTTTTCGCCAGACACATTCATGTGT 1908
DB |||||||
QY 1741 GTCTGAACACAGTGTCCATAGTGTGCTGAATATAAATCTCTGAAGAAAATGAACCAAGTGA 1800
DB |||||||
QY 1909 GTCTGAACACAGTGTCCATAGTGTGCTGAATATAAATCTCTGAAGAAAATGAACCAAGTGA 1968
DB |||||||
QY 1801 GGTTCGAAGAAAAGACACGATCAGAGTCGAGCGTGCAGTCGAGGGCCAGCTCGCCGGA 1860
DB |||||||
QY 1969 GGTTCGAAGAAAAGACACGATCAGAAAGTCGAGCGTGCAGTCGAGGGCCAGCTCGCCGGA 2028
DB |||||||
QY 1861 GGGTGTCTACAGCGGATGCTCTCGGCTTTCAGGAGCGCTCACCGGGAACCGAGATGACTT 1920
DB |||||||
QY 2029 GGGTGTCTACAGTGGATGTTTCTCGGCTTTCAGGAGCGCTCACCGGGAACCGAGATGACTT 2088
DB |||||||
QY 1921 ACCTTAAATGTGGGGCTGCAGTGTCTGCCAGATGTCGCCAGATGTCGAGGAGTTGACACAG 1980
DB |||||||
QY 2089 ACCTTAAATGTGGGGCTGCAGTGTCTGCCAGATGTCGCCAGATGTCGAGGAGTTGACACAG 2148
DB |||||||
QY 1981 GTGTAGCATAAAGTCAAGTCGTCTAACTTAAGATGCTCAGAGTTGTTTGTGGACTTCGC 2040
DB |||||||
QY 2149 GTGTAGCATAAAGTCAAGTCGTCTAACTTAAGATGCTCAGAGTTGTTTGTGGACTTCGC 2208
DB |||||||
QY 2041 TGTCTTCCCCCAAGAGCTGAAATGCTAAGCTACTTAAAAAGGATGCAAAAGCTTTGGTGT 2100
DB |||||||
QY 2209 TGTCTTCCCCCAAGAGCTGAAATGCTAAGCTACTTAAAAAGGATGCAAAAGCTTTGGTGT 2268
DB |||||||
QY 2101 GTGTAGTAAACAGAACCCCTGCTCTGCTGCTCAGGATGCTGCGGTTGGAGTGGAA 2160
DB |||||||
QY 2269 GTGTAGTAAACAGAACCCCTGCTCTGCTGCTCAGGAAATGCTGCGGTTGGAGTGGAA 2328
DB |||||||
QY 2161 ACAGAGCGCTGGAAATGATTGCTCGCCAGGTACCGAGAGAGAGCAGCTTTTAGGAGCTGGT 2220
DB |||||||
QY 2329 ACAGAGCGCTGGAAATGATTGCTCGCCAGGTACCGAGAGAGAGCAGCTTTTAGGAGCTGGT 2388
DB |||||||
QY 2221 TCCTGTAAACATTAATTAATTCGTCCTCCCAAGTGTGGTGTGGCAATGGAAGTGTAGCCCTTACT 2280
DB |||||||
QY 2389 TCCTGTAAACATTAATTAATTCGTCCTCCCAAGTGTGGTGTGGCAATGGAAGTGTAGCCCTTACT 2448
DB |||||||
QY 2281 TGAATGTATAGTGTAGATTTTAAACAGCAGGTTCTATATTATTATTATTATTATTATTATTATT 2340
DB |||||||
QY 2449 TGAATGTATAGTGTAGATTTTAAACAGCAGGTTCTATATTATTATTATTATTATTATTATTATT 2508
DB |||||||
QY 2341 TTTGGGATTAACCTCTTTTCATATGTTTGTCTGTACATAAAATACATAGCTATGTTAA 2400
DB |||||||
QY 2509 TTTGGGATTAACCTCTTTTCATATGTTTGTCTGTACATAAAATACATAGCTATGTTAA 2568
DB |||||||
QY 2401 GAGGCTTTAAAGTTTAAAACTTCAACCATGCTTGAGTATAGCAATTTTCATGCAATTTAA 2460
DB |||||||
QY 2569 GAGGCTTTAAAGTTTAAAACTTCAACCATGCTTGAGTATAGCAATTTTCATGCAATTTAA 2628
DB |||||||
QY 2461 AATGTTTTTCAGTGGCATGGTTTACAGAGGTTAGGACCACTGCCACATGACAGTTAAGA 2520
DB |||||||
QY 2629 AATGTTTTTCAGTGGCATGGTTTACAGAGGTTAGGACCACTGCCACATGACAGTTAAGA 2688
DB |||||||
QY 2521 CTTTATTTTTTAAAGCATCTCGGCAATAAAAATCAAAAGCCCTTTCATAAGCTGAGTTTCAAG 2580
DB |||||||
QY 2689 CTTTATTTTTTAAAGCATCTCGGCAATAAAAATCAAAAGCCCTTTCATAAGCTGAGTTTCAAG 2748
DB |||||||
QY 2581 ATAACTAGAACTACTAAACGTTTACATTTTTTTCAGATTTTTTAAAGCAATTTGATTTTTTAT 2640
DB |||||||
QY 2749 ATAACTAGAACTACTAAACGTTTACATTTTTTTCAGATTTTTTAAAGCAATTTGATTTTTTAT 2808
DB |||||||
QY 2641 ATATGCAATGTTATATTTTCTAAGAGGAATATTGATTATGGAGTAATGGGG 2692
DB |||||||
QY 2809 ATATGCAATGTTATATTTCTAAGAGGAATATTGATTATGGAGTAATGGGG 2860
DB |||||||

RESULT 5
AF532970
LOCUS Rattus norvegicus afadin- and alpha-actinin-binding protein ADIP
DEFINITION mRNA, complete cds.
AF532970 3195 bp mRNA linear ROD 03-FEB-2003

AF532970	AF532970.1	GI:27451840		QY	362	AATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTGCTGCTTACGCGGAAGAACCTG	421
VERSION				Db	361		420
KEYWORDS						AGTAGTGTGCTTCTGAACTGCAATGAGCTGCTTGTGTTCTTACGCGGAAGAACCTC	
SOURCE	Rattus norvegicus (Norway rat)			QY	422	CTGGCCAGGAGAGCGTGGAGACACAGAACTTGAAGCTGGGAGTGACATGACACACCTG	481
ORGANISM	Rattus norvegicus			Db	421		480
REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Rattus.			QY	482	CTGGCCAGGAAAGGCTGGAGACACAGAACTTGAAGCTGGGAGTGACATGACACACCTG	541
AUTHORS	1 (bases 1 to 3195)			Db	481		540
	Asada, M., Irie, K., Morimoto, K., Yamada, A., Ikeda, W., Takeuchi, M. and Takai, Y.			QY	542	CAGAGCTCTACGCCAAACTTAAAGGAGCAGTTGGAACGCTCAGCGCGGAGATGATCGGG	601
TITLE	ADIP, a Novel Afadin- and alpha-Actinin-Binding Protein Localized at Cell-Cell Adherens Junctions			Db	541		600
JOURNAL	J. Biol. Chem. 278 (6), 4103-4111 (2003)			QY	602	CAGAGCTCTACGCCAAACTTAAAGGAGCAGTTGGAACGCTCAGCGCGGAGATGATCGGG	661
PUBMED	12416711			Db	601		660
REFERENCE	2 (bases 1 to 3195)			QY	662	CTTCAAGAGAGACAGCAGCTGCAGTGCAGAAACAGAGATTTTCATCAGCTCTCTGAAG	721
AUTHORS	Asada, M., Irie, K. and Takai, Y.			Db	661		720
TITLE	Direct Submission			QY	722	AACGAGAAGAAGAGGTACAAAATATCAAAATATCATAGCCATGGATGTTTTAAATATATGTTGGGTGA	781
JOURNAL	Submitted (26-JUL-2002) Department of Molecular Biology and Biochemistry, Osaka University Graduate School of Medicine/Faculty of Medicine, Yamada-oka 2-2, Suita 565-0871, Japan			Db	721		780
FEATURES	Location/Qualifiers			QY	782	CTGATGATGTAAGAGAGCGTGAATATATAAAGCTAAAGAGCGCTCGCATCAG	
source	1..3195			Db	781		
	/organism="Rattus norvegicus"			QY	841	GCTGATGCAACAGAGGCTCATGGAGGACTCACAAAACAGAAAGCAGCAAGATGAAGATGAG	841
	/product="afadin- and alpha-actinin-binding protein ADIP"			Db	840		
	/protein_id="AAO15016.1"			QY	842	GTGATGGAAGCGAGGCTCTGGAGGACTGATAAAACAGAAAGCAGCAAGATGAAGATGAA	901
CDS	/db_xref="taxon:10116"			Db	841		
	79..1920			QY	902	ATGTACAAAAATCTGTTGAAATGATTTAGTACCGCCAGAAAGCAGATCCTGATGGAGAAC	901
	/codon_start=1			Db	901		
	/translation="MGDMWTVDVLPCTENKNSLYTSETKMSPLSYQQVLSATP			QY	962	ATGTACAAAAATCTGTTGAAATGATTTAGTACCGCCAGAAAGCAGATCCTGATGGAGAAC	901
	LSKNVGFSAFCTGENTEQISYLDQELTTGFPSPLYEESKSEAKELSVALLNC			Db	961		
	MNELLVQRNKLQAESVETQNLKLGSDHQLQSCYAKLKEQLSEARREMSLOERDR			QY	1021	CGGAGCTGGAAGAAGGCTCTCAGCAGATGAAGAAGAGATGATCTCTCTCTCTCTCTCT	960
	QLOQVRNLHOLLNKEVEVKLONIIASRATQYNDVVRKEREYKLRHLHOLVNR			Db	1020		
	KKDNIAVDLVNIVGRVDGRKGRSWRTDKTEARNEDEMYKILLNDYEVYRQKILLNAE			QY	1082	CAGAAAGAAAGCCAGGAAAGAGCAGACGACGCGCACAGGCTGTTGCTATCTCCGAT	1021
	LKVLQNMKEWISLPQKKPRERAEDSTGVVISVEDDAGELSRDGVNSLSCDT			Db	1081		
	VREQLTNSIRKQWLKSHVEKLQASVHSEGFHEEDVSRQDHEQTEKLEJLEI			QY	1142	CAAAAGAAAGAACCCAGAGAAAGAGCAGACACGCGACATGTTGTCTATCTCCGAT	1020
	RCKEMIQAQOQLLOQALATACDDDTLSLRDCYLLEEKERLEKSWLFPKQKQNFRE			Db	1141		
	RRSFEARAILGLRKAPEERASVWQFLNMTTFDHQNSENVKLSFSAFGSGSDPON			QY	1202	GTAGAGAGCAGCTGCACAAACAGCATCAGGAAAACAGTGGAGAAATTTTGAAGAAGTCATGA	1141
	LIVHPRPKPHSVANGVPACTSKLAKSLTPSPDFCPSRSCVSEHSPVSALTPTPE			Db	1201		
	ETKNEVGRESTDQKWSVVSFSPSSREGCYGCGSSAYTSSHVRRDLP"			QY	1261	GTAGAGAGCAGCTGCACAAACAGCATCAGGAAAACAGTGGAGAAATTTTGAAGAAGTCATGA	1141
ORIGIN				Db	1260		
	Query Match 81.3%; Score 2189.4; DB 6; Length 3195;			QY	1321	GTAGAGAGCAGCTGCACAAACAGCATCAGGAAAACAGTGGAGAAATTTTGAAGAAGTCATGA	1141
	Best Local Similarity 90.1%; Pred. No. 0;			Db	1320		
	Matches 2427; Conservative 0; Mismatches 247; Indels 20; Gaps 7;			QY	1381	GTAGAGAGCAGCTGCACAAACAGCATCAGGAAAACAGTGGAGAAATTTTGAAGAAGTCATGA	1141
				Db	1380		
QY	2	GTAGAGAGTACAGGAGCTGTTGTAAGCGTGGCAGCAGCTAGCGCGCTCTCAGGTATC	61				
Db	1	GTAGAGAGTACAGGAGCTGTTGTAAGCGTGGCAGCAGCTAGCGCGCTCTCAGGTATC	60				
QY	62	CTGGCTCTGGAACTTGCTATGGAGATTTGGATGCTGACAGATCCAGATTTCTGTGTACA	121				
Db	61	CTGGCTCTGGAGCTGGCTATGGAGATTTGGATGCTGTTACAGATCCAGTCTGTGTACA	120				
QY	122	GAAGAAACAAATCTCTCAATATACCTCAGAAACAAAGATGCTCCGTCAGATTTGTAC	181				
Db	121	GAAGAAACAAATCTCTCAATATACCTCAGAAACAAAGATGCTCCGTCAGATTTGTAC	180				
QY	182	TCCAGCAAGTCTGTGCTCTTACGTACCTTTATCCAAAACAGTGGTGTGTTTGGT	241				
Db	181	TCCAGCAAGTCTGTGCTCTTACGTACCTTTATCCAAAACAGTGGTGTGTTTGGT	240				
QY	242	GCTCTTCTCAGAGAGAGAAATTTGAACAAAGTATTTCTCTTGTATCAGGAGCTGACC	301				
Db	241	GCTCTTCTCAGAGAGAGAAATTTGAACAAAGTATTTCTCTTGTATCAGGAGCTGACC	300				
QY	302	ACCTTCGGGTTTCTCTCTTGTATGAAGATTTCAAAAGTAAAGAGGCAAGAGAGAAATTA	361				
Db	301	ACCTTCGGGTTTCTCTCTTGTATGAAGATTTCAAAAGTAAAGAGGCAAGAGAGAAATTA	360				

D	b	1441	ACAGAAGCTGCTATTTCGCTTGGGTTCGAGAGAAAGCGTTTGAGGAAGACGAGCCAGC	1500
Q	y	1502	TGGGTAAAGCAGCAGTTTTTAACATCACGNACTTTGCACCACCAACTCAGAAAATGTG	1561
D	b	1501	TGGGTGAAGCAGCAGTTTTTAACATCACGACCTTTGATCACCAGAACTCAGAAAAATGTG	1560
Q	y	1562	AACCTTTTCAGTGCCTTCTCAGGAAGTTCTGATCCAGACAATCTTATAGTCCACTCACGG	1621
D	b	1561	AACTTTTCAGTGCCTTTTCAGGAAGTTCTGATCCAGACAATCTTATAGTCCAACCAGG	1620
Q	y	1622	CCACGCCAAGAAGCTACACAGTGTGGCTAAATGGGGTCCAGCTTGCACATCAAACATG	1681
D	b	1621	CCACGCCAAGAAGCCACACAGTGTGGCTAAATGGGGTCCAGCTTGCACATCCAACTG	1680
Q	y	1682	ACTAAATCTCTTCTGTGCTCACCTTCTACTTTCAGACTTTCGCCAGACACATTCATGTG	1741
D	b	1681	GCTAAGTCTCTTCCGACTCAC-----CTTCAGACTTCTGCCCGTCTCGCTCATGTGTG	1734
Q	y	1742	TCTGAACACAGTTCCATCTCAGTGTGCTGAAATATACTCTGAAGAAAGTAAACCAAGTGAG	1801
D	b	1735	TCTGAGCACAGTCCCGTTCAGTGCCTGACTGTGACTCTCTGAAGAACCAACCGAATGAG	1794
Q	y	1802	GTTGCAAGAGAAAGCACCGGATCGAAGTGGAGCGGTGCAGTCGAGGCCAGCTCGCGGGAG	1861
D	b	1795	GTTGGAAGAGAAAGTACCGGACCAGAAGTGAGCGTGGTGTCCAGACCAGCTCCCAGGAG	1854
Q	y	1862	GGGTGCTACAGCGGATGCTCCTCGGCCCTTCAGGAGCGCTCACGGGACCCAGATGACTTA	1921
D	b	1855	GGTTGCTACCGGTGGATGCTCCTCGGCCCTACACAAGACTCCCCGCTGGAAACGAGATGACTTA	1914
Q	y	1922	CCTTAAATGTCGGGGTGCAGTGTGTTTCCAGATGTCGCTAGAGAGTTGATTTGGACTTCGCT	1981
D	b	1915	CCATAGGTGCTCGGGCTGCAGCGCTGT--CCTGAGTGCATGAGAGGAATTGACACGGGG	1972
Q	y	1982	TGTAGCATAAAGTCAGTCGTCTAACTTAAGATGCTCAGAGTGTGTTGTTTGGACTTCGCT	2041
D	b	1973	TGTAGCATAAAGTCAGCCATCTACCGTAAAGATG-TCGGAGTTATTTGTTTGGACTTCCCA	2031
Q	y	2042	GTCCTCCCCAAGAGCTGAATGCTTAAGCTACTTAAAGGNTGCAAGCTTTGGTTGTG	2101
D	b	2032	GTCCTTCCCCAAGAGCTGAACGCT-----TTAGAGGATGCGAAAGCTTTGGCTGTG	2084
Q	y	2102	TGTTAGTAAACAGAAGCCCTGGCTCTGTGACTGCAGGAATGCAATGGCGTTTGGATGGAAA	2161
D	b	2085	TGTTAGTAAACAGAAGCCCTGGCTCTGTGAGTAAAGGAATGATGTGTTTGGTGGGAAA	2144
Q	y	2162	CAGAAGCGCTGGAATGATTCGCTC-GCCAGGTACCGAGAAGACACTTTTAGGGACTGGT	2220
D	b	2145	CAAAAGCACAGAAATGATTTTCTCTCCGGTACTGAGAATAGCACCTTTTAGGAGCTGAT	2204
Q	y	2221	TCCTGTAACACTTAANATTCGTCGCCAAGTGTGTTGGCAATGGGAAGTGTAGCCTTTACT	2280
D	b	2205	TCCTGTAACACTTAANATTTTGTGCCAAGTATGGTTGGCAATGGGAAGCTTTAGTCTTTACT	2264
Q	y	2281	TGAATGTATCTGTAGATTTTTTAACAAGCAGGTTCTATATTTATTATGTTTATGTGTGAT	2340
D	b	2265	TGAATGTACCTGTAGATTTTTTAACAAGCA-GTTCTATATTTATTAAGTTTATGTGTGAT	2323
Q	y	2341	TTTGGGATTAACCTCTTTTCATATGTTTGTGTCGTGTACATAAATATACATGCTATGTTAA	2400
D	b	2324	TTTTGGATTAACCTCTTTTCATATGTTTCTCGCCTGTACATAAATATACATGCTATGTTAA	2383
Q	y	2401	GAGGCTTTAAGGTTTAAAACTTCACACCATGCTTTGAGTATAGCAATTTTCATGCCAATTA	2460
D	b	2384	GAGGCTTTAAGGTTTAAAAATTTTCACACCATGCTTCGAGNATAGCAATTTTCATGCCAATTA	2443
Q	y	2461	AATGTTTTTCAGTGGCATGGCTTTTACAGA--GGTTAGGACCACTGCCACATGACAGTTAA	2518
D	b	2444	AATGTTTTTCAGTGGCATGGCTTTTACAGATGTGTTTAGGACCACTGCCACATGACAGTTAA	2503
Q	y	2519	GACTTTTATTTTAAAGCCATCTGGGCAATAAAAATTTCAAAGCCCTTTCATTAAGCTGAGTTC	2578

Db	2504	GATTTTATTTTTAAAGCAATTTGGGCAATAAAAAATTCAAAGCCACTTCATTAAGCTAAGTTC	2563
Qy	2579	AGATAACTAGAACTACTAAACGTTTACATTTTTTGAGATTTTTTAAAGCACTTGTATTTTATTTT	2638
Db	2564	AGATAGCTAAACTACTAACATTTTACATTTTGGAGATTTTAAAGCATTATTTTATTTT	2623
Qy	2639	ATATATGTGGAATGTTTATAATTTCTAAGAGGAATATTTGATTTATGAGTAATGGGG	2692
Db	2624	ATATATGTGACTGTGTATAATTTCTAAGAGGAATGTGGATTATGAAGCAATGGGG	2677
RESULT 6			
BC078687			
LOCUS	BC078687	2547 bp	mRNA linear ROD 09-DEC-2005
DEFINITION	Rattus norvegicus synovial sarcoma, X breakpoint 2 interacting protein, mRNA (cdna clone MGC:93056 IMAGE:7115887), complete cds.		
ACCESSION	BC078687		
VERSION	BC078687.1		
KEYWORDS	GI:50926757		
SOURCE	MGC.		
ORGANISM	Rattus norvegicus (Norway rat)		
REFERENCE	1 (bases 1 to 2547)		
AUTHORS	Strausberg, R.D., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, B., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, D.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Vallal, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalski, J., Smalusz, D.E., Schnarch, A., Schein, J.E., Jones, S.J. and Marra, M.A.		
CONSTRMT	Mammalian Gene Collection Program Team		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 2547)		
AUTHORS	NTH MGC Project		
CONSTRMT	Direct Submission		
TITLE	Submitted (02-AUG-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA		
JOURNAL	NTH-MGC Project URL: http://mgc.nci.nih.gov		
REMARK	Contact: MGC help desk		
COMMENT	Email: csapba@mail.nih.gov Tissue Procurement: Howard Jacobs cDNA Library Preparation: Express Genomics cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305-5051 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcd@paxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAC Plate: 183 Row: 1 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 31342223. Location/Qualifiers		

source	1..2547 /organism="Rattus norvegicus" /mol_type="mRNA" /db_xref="taxon:10116" /clone="MGC:93056 IMAGE:7115887" /tissue_type="Testis, rat (Brown Norway)" /clone_lib="NIH MGC_237" /lab_host="DH10B" /note="Vector: pExpress1" 1..2547 /gene="Sex2ip" /note="synonym: ADIP" /db_xref="GeneID:308023" /db_xref="RGD:708490" 366..2207 /gene="Sex2ip" /codon_start=1 /product="synovial sarcoma, X breakpoint 2 interacting protein" /protein_id="AAH78687.1" /db_xref="GI:50926758" /db_xref="GeneID:308023" /db_xref="RGD:708490" /translation="MGDMWTVDPVLTENKNLSOYTSETKMSPSLSYQOVLCSATP LSKNVGVFSAPCTGENTEQSTSYLDQELTTGPPSLYEESKREKRELSIVALLNC MNELLVLORKNLLAQESVETQNLKLGSDMDHLQSCYAKLKEQLERSRREMISLQERDR QLQCKRNHLHOLLKNEKEVEVKLQNIIASRATQYNHDKRKEREYNKLKERLHOLVWN RKDNIAMVMVLYGRVQKGRSWRTDTEARNEDMYKILLNDYEVKQKILLSENAE LKKVLQMKMISLLSPQKKPRERAEDSTGVVISEDDBAGELSDRGVKSLSCDT VREOLTNSIRKQWRILKSHVEKLDNQASKVHSEGHREDVLSRODHEOTKLELEIE RKEMIKAQOOLLQALACDDPTSLRLDCYLLEKERLKEWSLFEKQKNPFERE RRSTFEAIRGLERKAFERERASWVKQFLNMVTTDHQNSENVKLFSAFGSSDPDN LIVHPRQKQKSVANVGPACTSKLAKSLPTSPDFCSRCSSEHSPSVSALTVTPE ETKPEVNGRESTDQKSVSVSRPSREGCYGCGSSAYTSSHVERDLP"
gene	
CDS	
Db	767 GCAGAGCTGCTACGCCAAACTTTAAGGAACAGTGTGGAGGCTCTCAGCGAGAGATGATCAG 826
Qy	541 GCTTCAAGAGAGAGACAGCAGCTGCAGTGCAGGAACAGGAGTGTTCATCAGCTCCTGAA 600
Db	827 CTTTCAGAGAGAGACAGCAGCTACAGTGCAGGAACAGGAAATTTGCATCAGCTCCTGAA 886
Qy	601 GAATGAGAAAGATGAGGTACAAAAATACAAAATATCATAGCCAGCGGGCTACTCAGTA 660
Db	887 AAACGAGAAAGAGGTACAAAAATACAAAATATCATAGCCAGTCTGGGCTACTCAGTA 946
Qy	661 TAATCATGATGTGAAGAGGAGGAGCGTGAATATATTAAGCTTAAGGAGCGCTGCATCA 720
Db	947 TAATCATGATGTGAAGAGGAGGAGCGGAGTACAAATAAAGCTGAAGGAGCGCTGCATCA 1006
Qy	721 GCTCGTTATGAACAAGAGGATAAAAACATAGCCATGATGTTTAAATATATGTCGGTCG 780
Db	1007 GCTTGTATGAACAAGAGGATAAAAATATAGCCATGAGACGTTTAAATATAGCTGGCGG 1066
Qy	781 AGCTGATGGCAAA CGAGGCTCATGAGGACTGACAAAAACAGAACGACGAGATCTGATGAG 840
Db	1067 AGTGATGCAAAAGCGAGGCTCTCGAGGACTGATAAAAACAGAACGACGAGATCTGATGAG 1126
Qy	841 GATGTACAAAATCTGTTGAATGATTATAGTACCGCCAGAGCAGATCTCTGATCGAGAA 900
Db	1127 AATGTACAAAATCTGCTGAATGATTATAGTACCGCCAGAGCAGATCTCTGCTGGAGAA 1186
Qy	901 CGCGAGCTGAAGAAGGCTCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTCTCTCC 960
Db	1187 TCGGAGCTGAAGAAGGCTCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTCTCTCC 1246
Qy	961 TCAGAAAGAAAGCCCGAGGAAAGAGCAGAGGACGCGACAGGACACTGTTGCTATCTCCGA 1020
Db	1247 TCAAAAGAAAGAAACCCAGAGAAAGCAGCAGAGCAGCAGCACTGTTGTCATCTCCGA 1306
Qy	1021 TATAGAGATGACTCTGGGAACTTGAGCAGAGACAGCGTGTGGGCGCTTCTCTGTGACAC 1080
Db	1307 TGTAGAAGACACGCTGGGGAGCTGAGCAGAGATGGTGTGTGGAGCGCTTCTCTGTGACAC 1366
Qy	1081 TGTGAGAGAGCAGCTGACAAAACAGCATCAGGAACAGTGGAGAAATTTTGAAGGTCATGT 1140
Db	1367 TGTGAGGAGAGCAGCTTACAAAACAGCATCAGAGAGCAGTGGAGAAATTTGAAAAGCAGTGT 1426
Qy	1141 AGAAAACTCGATAACCAAGCTTCGAAAGTACACTCAGAGGCGCTTAAATAGGAGGACGT 1200
Db	1427 GGAATACTTGATAACCAAGCTTCAAAGTACACTCAGAGGCGCTTCTCATGAAGAGGACGT 1486
Qy	1201 CATCTCAGCAAGACCATGAGCAGAGACTGAGAACTGAGAGCTGGAGATTTGAGCGGTG 1260
Db	1487 CATCTCAGCAAGACCATGAGCAGAGACTGAGAACTGAGAGCTGGAGATTTGAGCGGTG 1546
Qy	1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACACGCTGTGA 1320
Db	1547 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCTGCGTGTGA 1606
Qy	1321 TGATGACACCACTCTCTGTTGCGAGACTGTTACTGCTGGAAGAAAAAGGAAACGCTTAA 1380
Db	1607 TGACGACACCACTCTCTGTTGCGAGACTGTTACTGCTGGAAGAAAAAGGAAACGCTTAA 1666
Qy	1381 AGAGAGTGGACCTCTTTTAAAGGACAAAAGAAATTTTCAGAGAGAAAGGCGAAGCTT 1440
Db	1667 AGAAGTGGTCTCTCTTTTAAAGGACAAAAGAAATTTTCAGAGAGAAAGGCGAAGCTT 1726
Qy	1441 TACAGAGCTGCCATTCGATTGGGTTGGAGAGAAAGCGCTTTCAGAGAGAGAGCGGCGAG 1500
Db	1727 TACAGAGCTGCTATTCGTTGGGTTGGAGAGAAAGCGCTTTCAGAGAGAGAGCGGCGAG 1786
Qy	1501 CTGGGTAAAGCAGCAGTGTTTAAACATGACGAACTTTGACCAACAGAACTCAGAAAAATGT 1560
Db	1787 CTGGGTAAAGCAGCAGTGTTTTAAACATGACGAACTTTGATCACCAGAACTCAGAAAAATGT 1846
Qy	1561 GAAACTTTTCACTGCGCTTCTCAGGAAGTTCGTATCCAGACATCTTATAGTCCACTCAGC 1620

ORIGIN

Query Match	67.2%; Score 1810.2; DB 6; Length 2547;
Best Local Similarity	89.3%; Pred. No. 0;
Matches 2011; Conservative	0; Mismatches 223; Indels 17; Gaps 5;
Qy	1 CGTAGGAGATGACAGGAGCTGTTGTAAGCGCTGCGCAGCAGCTGAGCGCGCTCTCAGGTAT 60
Db	287 CGTAGGAGATGACAGGAGCTGTTGTGCATGCCCGCAGCAGTGAAGTCTCAGGGAC 346
Qy	61 CCTGCTCTGGAACCTTGCTATGGAGATGATGATGATGATGATGATGATGATGATGATGAT 120
Db	347 CCTGCTCTGGAACCTTGCTATGGAGATGATGATGATGATGATGATGATGATGATGATGAT 406
Qy	121 AGAAACAAATCTCTCAATATACCTCAGAACAAAGATGCTCCGTCAGTTGTA 180
Db	407 AGAAACAAATCTCTCTCAATATACCTCAGAACAAAGATGCTCCGTCAGTTGTA 466
Qy	181 CTCCAGCAAGTCTGCTCTTCACTACCTTTATCCAAAACGTCATGCTGTTTCGG 240
Db	467 CTCCAGCAAGTCTGCTCTTCACTACCTTTATCCAAAACGTCATGCTGTTTCAG 526
Qy	241 TGTCTTCTGCAAGAGAGAACTTGAACAAAGTATTTCTTATCTGATCAGGAGCTGAC 300
Db	527 TGCCTTCTGCAAGAGAGAACTTGAACAAAGTATTTCTTATCTGATCAGGAACTGAC 586
Qy	301 CACCTTCGGGTTTCTCTCTGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 360
Db	587 TACCTTCGGGTTTCTCTCTGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 646
Qy	361 AAATATAGTCTGCTGTTGAACTGATGAACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db	647 AAGTATAGTCTGCTGTTGAACTGATGAACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 706
Qy	421 GCTGCCCAAGAGCGTGGAGACACAGAACTTGAAGCTGGGCGAGTGAATGAGCCACCT 480
Db	707 CTTGCCCAAGAGCGTGGAGACACAGAACTTGAAGCTGGGCGAGTGAATGAGCCACCT 766
Qy	481 GCAGAGCTGTACGCCAACTTAAGGAGCAGTTTGAAGAGCTCCAGGCGGGAGATGATCGG 540

ORIGIN	EBIKPNQVGECTNQKWSVASRPGSQEGCYSGCSLSYNSHVEKDLP"		Db
Query Match	63.2%;	Score 1701.4;	DB 5; Length 3220;
Best Local Similarity	80.7%;	Pred. No. 0;	
Matches 2214;	Conservative 0;	Mismatches 471;	Indels 60; Gaps 17;
QY	1	CCTAGGAGAGTACAGGAGCTGTGTAAAGCGTCGACGACTAGAGCGCCCTCTCAGGTAT	60
Db	73	CATAGAAGAGTGACAGCAGTACACTAAATGTTTAACTGCTGAAGTGTCTCAGGTAT	132
QY	61	CCTGGCTCTGGAACTTGTCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCGTGTAC	120
Db	133	CCTGGCTCTAGAGATTGCTATGGGAGATTGGATGACTGTTCAGATCCAGGTCTGTCTTC	192
QY	121	AGAAAAAATAATCTCTCTCAATATACCTCAGAAAAAAGATGTCTCCGTCAGTTTGTGA	180
Db	193	AGAAAGCAAAATCTATCTCTCAATATACCTCAGAAAAAAGATGTCTCCATCAAGTTTATA	252
QY	181	CTCCAGCAAGTCTGTGCTCTTACGTACCTTTATCCAAAAACGTGCGATGGTGTTCGG	240
Db	253	CTCAGCAAGTGTGTATGTTCTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTCAG	312
QY	241	TGCTCTTCGCACAGGAGAAATTTGACAAAGATATTTCCATCTTGATCAGAGCTGAC	300
Db	313	TGCTCTTCGCACAGGAGAAATTTGACAAAGATATTTCCATCTTGATCAGGAAATGAC	372
QY	301	CACCTTCGGGTTTCCCTCTGTATGAAGAAATCCAAAAAGTAAAGGCAAGAGAGAAAT	360
Db	373	TACTTTTGGTTTCCCTCATTATATGAAGAAATCCAAAGGTAAAGAGACAAAGAGAGATT	432
QY	361	AAATATAGTCGTGTCTGAACTGTATGAAACGAGTGTCTGTCGTTTCAGCGGAAAGAACT	420
Db	433	AAATATAGTGTGTACTAAATTTGTATGAATGAGTGTCTGTCGTTTCAGCGGAAAGAACT	492
QY	421	GCTGCCCAGGAGAGCGTGGACACAGAACTTTGAAGCTTGGCAGTGCATGACAGCACT	480
Db	493	TCTAGCTCAGGAAATGTGGAGACAGAAATTTGAAGCTTGGGAGTGGAGTATGGAACATCT	552
QY	481	GCAGAGCTGTACGCCAAATTTAAGGAGCAGTTGGAACAGTCCAGCGGGAGATGATCGG	540
Db	553	ACAGAGCTGTACTCAAAACTTTAAGGAACACTGGAAACCTCCAGGAGGGAATGATTTGG	612
QY	541	GCTTCAAGAGAGACAGCGAGCTCAGTGCAGAAACAGGAGTTTGGCATCAGCTCCTGAA	600
Db	613	GCTTCAGGAAGAGACAGACAGTTTCAATGTGAAGAACAGGAATTTGCATCAGCTACTAAA	672
QY	601	GAAACAGAAAGATGAGGTACAAAAATTTACAAATATCATAGCCAGCGGCTACTCAGTA	660
Db	673	GAAACAGAAAGATGAGGTGCAAAAAATTTACAAATATCATAGCTCAGCTACTCAGTA	732
QY	661	TAATCATGATGTGAAGAGAGGAGCGTGAATATATAAGCTTAAAGGAGCGCTGCATCA	720
Db	733	TAATCATGATGAAGAGAGAGAGCGGTGAATATATAAGCTTAAAGGAGCGCTGCATCA	792
QY	721	GCTGTTATGAACAGAGAGGATAAAAAATAGCCATGATGTTTAAATATATGTTGGGTCG	780
Db	793	ACTTGTATGAACAGAGAGGATAAAAAATAGCTATGGACATTTTGAATTTATGTCGGGAG	852
QY	781	AGCTGATCGCAACGAGGCTCATGGAGCTCAGCAAAACAGAGCCAGGAATGAGATGA	840
Db	853	AGCTGATCGCAACGAGGCTCCTGGAGCTTGGTAAAACTGAAGCCAGGAATGAGATGA	912
QY	841	GATGTACAAAAATTTCTGTTGAATGATTATGAGTACCGCCAGAGCAGAGATCTCTGATGGAAA	900
Db	913	AATGTATAAATTTCTCTTGAATGATTATGAATATCGTCAGAAACAAATCTTAATGAANA	972
QY	901	CGCGAGCTGAAGAGGTCCTCCAGCAGATGAAGAGAGATGATCTCTCTCTCTCTCTCC	960
Db	973	TGCAGAACTTAAGAGGTTCTTCAACAAATGAAGAAAGGAATGATTTCTCTCTCTCTCC	1032
QY	961	TCAGAGAGAGCCACGGGAAAGAGCAGAGGACGGCACAGGCACTGTGTATCTCCGA	1020

1033	CCAAAAGAGAAACCTAGAGAAAGAGTAGATGATAGTACAGAACTGT---	TATTCCGA	1089
1021	TATAGAAGATGACTCTGGGGAACCTAGAGAGACAGCGTGTGGGCTTTCTCTGTGACAC	1080	
1090	TGTTCAAGAGAGATGCGCGGGAACCTAAGCAGAGAGAGTAGTGTGGACCTTTCTCTGTGAAC	1149	
1081	TGTGAGAGAGCAGCTGACAAACAGCATACAGGAAAACAGTGGGAAATTTTGAAGAGTCATGT	1140	
1150	TGTGAGAGAGCAGCTTACAAACAGCATCAGAAAAACAGTGGAGAAATTTTGAAGAGTCATGT	1209	
1141	AGAAAACTCGATTAACCAAGCTTCCAGGTACACTCAGAGGCCCTTTATAGAGAGAGACGT	1200	
1210	AGAAAGCTTGATAACCAAGTTTCAAGGTACACCTGGAAGGTTTAAATGATGAAGATGT	1269	
1201	CATCTCAGCAACAAGACCATGAGCAAGAGACTGAGAAAACTGGAGCTGGAGATTGAGCGGTG	1260	
1270	AATCTCAGCAACAAGACCATGAGCAAGAACTGAAAACTCGAGTTAGAAATTTACGACGTG	1329	
1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTTTACAGCAGCAGCTGGCCACACGCTGTGA	1320	
1330	TAAAGAAATGATTAATACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCATATGA	1389	
1321	TGATGACACCACTCAGCTGTTGCGAGACTGTTACTTCTGGAGAAAAAGGAAGCCTTAA	1380	
1390	TGATGATACCACTTCACTATTACGAGACTGTTATTGTTGGAGAAAAAGGAAGCCTTCAA	1449	
1381	AGAGAGTGGACCCCTTTTAAAGCAAAAAAAGAAATTTGAGAGAGAAAAAGGCGAAGCTT	1440	
1450	AGAGAAATGCTCCCTTTTAAAGCAGCAAAAAAGAAATTTGAGAGGAGAGACGAAAGCTT	1509	
1441	TACAGAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGTTTGAAGAGAGCGAGCCAG	1500	
1510	TACAGAGCCGCTATTTCGCTGGGATTGGAGAGAAAGCAATTTGAAGAGAAAGAGCCAG	1569	
1501	CTGGTAAAGCAGCAGTTTAAACATACGAACTTTGACCCACAGAACTCAGAAATGT	1560	
1570	TTGGTTAAAGCAGCAGTTTCTAAATATGACTACCTTTGACCACCAAGAACTCAGAAATGT	1629	
1561	GAACCTTTTCAGTGCCTTTCTCAGGAAGTTCTGATCCAGACAATCTTATAGTCCATCAGC	1620	
1630	GAACCTTTTCAGTGCCTTTCTCAGGAAGTTCTGATTTGGACAACTCTTATAGTGCATCGAG	1689	
1621	GCCACGGCAAAAGAGTACAACAGTGTGGCTTAATGGGTGCGAGTTGACATCAAAACT	1680	
1690	GCAGCGCAAAAGAGCCTCAACAGTGTCTTAATGGGTCTCCAGTTTGCATGCTTAAACT	1749	
1681	GACTAAATCTCTCTGCTCCTCAGCTTCTACCTTACAGCTTTGCGCCAGACATCTCATGTGT	1740	
1750	TACTAAATCTCTCTGCTTCCCTTCCACTTCCAGCTTTTGGCCAGACACGTTCTCTGCAT	1809	
1741	GTCTGAACACAGTTCCTCAGTGTGCTGAATATAACTCCTCGAAGAAAAGTAAACCAAGTGA	1800	
1810	ATCTGAACATAGTTCNATCAATGACTGTAATATAACTGCTGAGAAATTTAAACCAATCA	1869	
1801	GGTTGCAAGAGAAAGCAGGATCAAGAGTGAGCGGTGAGTGCAGGCCCAAGCTCCGGGA	1860	
1870	GGTTGGAGAGAAATGTAACAAATCAAAATGAGTGTGGCGTCAAGACCTGGATCACAGGA	1929	
1861	GGGTGCTTACAGCGGATCTCCTCGGCTTCCAGGCGCTCAGCGGACCGGAGATGACTT	1920	
1930	AGGTGCTTATAGTGGATGCTCTTTGAGCTACAAAATTTCTCATGTAGAAAAAGAGTACTT	1989	
1921	ACCTTAAATGTGCGGGTGCAGTGTCTTCCAGATGTGCGCTAGAG-----A	1969	
1990	ACCTTAGACATGTGAGCTGGAAATTTTTTTCATTATGTGTTCATCAAGTTTACATCTAA	2049	
1970	GTTGACACAGGCTGTAGCATAAAGTCAGTCTCT-----TAACCTTAAGATGCTCAGAGTTG	2024	
2050	GTTGAAACAGGCTGTGTATAAAGTCAGTTATCTCTAATAACTTAAAGATGCTCTGAGTTG	2109	
2025	TTTCTTTGGACTTCTCTCTTCCCAAGAGCTGAAATGCT--AAGTACTTAAAGGA	2083	
2110	TTTGTGAGACTTCCCTGTCTTCCCAAGAGTTGAAATCTTAAATCTATTAAAAAGGA	2169	

Qy	2084	TGCAAGCTTTGG-----TTGTGTTGTAGTAACAGAAGCCCTGGCTCTGTGACTGCAGGA	2139
Db	2170	TATAAAAGCTTTGGATATGTAATTTTGTAGTAACAGAAGCATCTGGTTCTGTGAATAAAGGA	2229
Qy	2140	ATGCATGCG- GTTTGGATGGAAAACAGAGCGCTCGAATGA--TTGCCCTCGCCAGGTACCG	2196
Db	2230	ATGTATAGATGTTTGGATGGAAAACAAAGCACTAGACTGAGTTTCTCTTATAGSTATTA	2289
Qy	2197	AGAAGAGACATTTTAGGACCTGGTTCCTGTGTAACATTTAAATATTCGTCGCCAAGTGTGGTT	2256
Db	2290	AAAAATAGCACATTTTAGGAAACTGATTAATGTGTAATAATGTTTAATTTTGTCTCAATATAGTT	2349
Qy	2257	GGCAATGGAAGTGTAGCCTTTACTTGAATGTATCTGTAGATTTTAAACAAGCAGGTTTC	2316
Db	2350	GGCAATGGAAGTGTAGCCTTTACTTGAATGTATCTGTAGATTTTAAACAAGCAGGTTTC	2409
Qy	2317	TATATTTATTTAGTTAGTGTGATTTTGGGATTAACCTCTTTTCATATGTTTT-----	2367
Db	2410	TATATTTATTTAGTTAGTGTG- GTTTGAAATTAACCTCTTTTCATATGTTTTAAATAAAGT	2468
Qy	2368	-----GTGCTGTACATAAATATATACATGACTATGTTTAAGAGGCTTTTAAGGTTT	2415
Db	2469	GAAATTTTATGTATGTTTGTACATAGATACATGATTTATGTTAAGAGGCTTTTAAGATTT	2528
Qy	2416	AAAAACTTTCACACCATGCTTCAGTATAGCATTTTCATGCCAATTTAAATGTTTTTCAGTGGC	2475
Db	2529	AAAAGTTTCACA-CAACCATAGTATAGTATTTTCATGCC-AGTAAATTTTTTTTAGTGGT	2586
Qy	2476	ATGTTGTTTACAGAGG--TTAGGCCACTGCCACATGACAGTTAAGACTTTTATTTTAAAG	2533
Db	2587	ATTCGTTTACAGATGTATTAGGACCATTTGATGCATTTACATTTAAGA-ATTCTCTTTAAT	2645
Qy	2534	CCATCTGGGCAATAAAATTCAGGCCCTTCAT-AGCTGAGTTC---AGATAACTAGA	2589
Db	2646	ACATCTGGGCAATAAATTTGAAAGGTATTCCATGAAGCTGAGTCTTTAGATAATCAAC	2705
Qy	2590	ACTACTACGTTTACATTTTTCAGATTTTAAAGCATT-GTATTTTATTTTATATATGT-G	2647
Db	2706	ACTACTAACATTTACATTTTTCAGATTTTATGACATTAGATTTTATTTTGTATATGTAG	2765
Qy	2648	AATGTTATATTTCTAAGAGGAATATTGATTTATGGAGTAATGGGG	2692
Db	2766	AATATTATATTTTAAAAAGGACTATTGATGATAGAAATAGGG	2810
RESULT 8			
CQ722034			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Homo sapiens			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
Hominidae; Homo.			
1			
Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.			
Kites, such as nucleic acid arrays, comprising a majority of			
humanexons or transcripts, for detecting expression and other uses			
thereof			
Patent: WO 02068579-A 7968 06-SEP-2002;			
PE Corporation			
Location (NY) (US)			
1..5835			
/organism="Homo sapiens"			
/mol_type="unassigned DNA"			
/db_xref="taxon:9606"			
JOURNAL			
AUTHORS			
TITLE			
FEATURES			
source			
ORIGIN			
Query Match			

Best Local Similarity 80.7%; Pred. No. 0;			
Matches 2214; Conservative 0; Mismatches 471; Indels 60; Gaps 17;			
Qy	1	CGTAGAGAGTGACAGAGAGCTGTTGTAAGCGTGCAGACACTGAGCCGCCCTCCTCAGGTAT	60
Db	186	CATAGAAGAGTGACAGCAGCTAGACTAAATGTTTAACTGCTGAATACTAGTTCTCCTCAGGTAT	245
Qy	61	CCTGGCTCTCGAACTTCCTATGGAGATTGGAGCTGTGACAGATCCAGTTCCTGTGTAC	120
Db	246	CCTGGCTCTAGAGATTGCTATGGAGATTGGAGCTGTGACAGATCCAGTTCCTGTGTAC	305
Qy	121	AGAAAACAAAAAATCTCTCTCAATATACCTCAGAAAAAAGATGTCTCCGTCCAGTTGTGA	180
Db	306	AGAAAGCAAAACTATCTCTCAATATACCTCAGAAAAAAGATGTCTCCATCAAGTTTATA	365
Qy	181	CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTATCCAAAAACGTGCATGTTGTTTCGG	240
Db	366	CTCACAGCAAGTGTGTTCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTCAG	425
Qy	241	TGCTCTCTGCACAGGAGAGAAACATTGAAACAAAGTATTTCTCTATCTTGATCAGGAGCTGAC	300
Db	426	TGCTCTCTGCACAGAGATTAATTGAAAGAGTATCTCATATCTTGATCAGGAATTGAC	485
Qy	301	CACCTTCGGGTTTCTCTCTTGTATGAAGAAATCCAAAGTAAAGAGGCAAGAGAGAAATT	360
Db	486	TACTTTTGGTTTCTCTTCAATTATATGAAGAAATCCAAAGGTAAGAGACAAAGAGAGAGTT	545
Qy	361	AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTCGTCTTCAGCGGAAGAACCCT	420
Db	546	AAATATAGTAGTGATCTAAATTTGTATGAATGAGCTGCTGTGTCTTCAGCGGAAGAACCCT	605
Qy	421	GCTGGCCACGAGAGCGTGGAGACACAGAACTTTGAAGCTGGGAGTGACATGAGACCACT	480
Db	606	TCTAGCTCAGGAAAAATGTGGAGACACAGAAATTTGAAGCTGGAGTGTATGAGCACTCT	665
Qy	481	GCAGAGCTGTACGCCAAAATTTAAGAGAGAGTTGGAAAAAGCTTCAGCGGGAGAGATGTCGG	540
Db	666	ACAGAGCTGTACTCAAAAATTTAAGGAACAACTTGAAACCTCCAGAGGGGAAATGATTGG	725
Qy	541	GCCTCAAGAGAGACAGAGCGAGCTGCAGTGCAGAACAGAGAGTTTGCATCAGCTCCTGAA	600
Db	726	GCCTCAGGAAAGAGACAGACAGTTTCAATGTAAAGACAGGAATTTTGATCAGCTACTANA	785
Qy	601	GAATGAGAAAAGATGAGGTACAAAAATTTACAAAATATCATAGCCAGCGGGCTACTCAGTA	660
Db	786	GAATGAGAAAGATGAGGTGCAAAAATTTACAAAATATCATTTGCAAGTGCAGCTACTCAGTA	845
Qy	661	TAATCATGTGTGAAGAGGAGGAGCGTGAATATAATTAAGCTAAAGGAGCGCTGCATCA	720
Db	846	TAATCATGATATGAAGAGAAAAAGAGCGTGAATATAATTAACCTGAAGGAACGCTACATCA	905
Qy	721	GCTCGTTTATGAACAAGAGGATAAAAACATAGCCATGATGTTTTTAAATTTATGTGGGTGCG	780
Db	906	ACTTGTATTGAACAAGAAAGATTAAGAAAAATAGTATGGAACATTTTGAATTTATGTGGGAG	965
Qy	781	AGCTGATGGCAACGAGGCTCATGAGGAGCTGACAAAACAGAAAGCCAGGAATGAAGATGA	840
Db	966	AGCTGATGGAAAAAGAGGCTCTCGAGGAGCTGTTGTTAAACTGAAGCCAGGAATGAAGATGA	1025
Qy	841	GATGTACAAAATTTCTGTTGAATGATATAGTACCGCCAGAGAGAGATCCTGTATGGAGAA	900
Db	1026	AATGTATAAAATTTCTTCTGAAATGATTAATATCGTCAGAAAAAATCCTTAATGGAAAAA	1085
Qy	901	CGCGGAGCTGAAGAAGGTCTCTCCAGCAGATGAAGAAGAGATGATCTCTCTCTCTCTCC	960
Db	1086	TGCAGAACTTAAGAAGGTTCTTCAACAAATGAANAAGGAATGATTTCTCTCTCTCTCC	1145
Qy	961	TCAGAAGAAGAGCCAGGGAAGAGCAGAGGACCGCAAGGCACTGTTGTATCTCCGA	1020
Db	1146	CCAAAAGAAGAAACCTTAGAGAAAGAGTAGATGATAGTACAGGAACTGT---TATTTCGA	1202
Qy	1021	TATAGAAGATGACTCTCTGGGAACTGAGCAGAGACAGCGTGTGGGCGCTTTCTCTGTGACAC	1080

Db 1203 TGTGTAAGAAGATGCGGGGAACCTAAGCAGAGAGATATGTGGACCTTTCCTGTGAAC 1262
Qy 1081 TGTGAGAGAGCAGCTGACAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAAGTCATGT 1140
Db 1263 TGTGAGAGAGCAGCTTACAAACAGCATCAGAAACAGTGGAGAAATTTTGAAAGTCATGT 1322
Qy 1141 AGAAAACTCGATAAACCAAGCTTCGAAGGTGACACTCAGAGGGCCTTAATGAGGAGACGT 1200
Db 1323 AGAAAACTTGTATACCAAGTTTCAAAGGTGACACCTGGGAAGGTTTTTAATGATGAAGATGT 1382
Qy 1201 CATCTCAGCAAGACCATGAGCAAGAGACTGAGAAACTGGAGCTGGAGATTTGAGCGGTG 1260
Db 1383 AATCTCAGCAAGACCATGAGCAAGAACTGAAAACTCGAGTTAGAAATTCAGCAGTG 1442
Qy 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGCTGTA 1320
Db 1443 TAAAGAAATGATTAATACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCATATGA 1502
Qy 1321 TGATGACACCACTCACTCTGCGAGACTGTTACTTGTCTGGAAGAAAGGAAAGCGCTTAA 1380
Db 1503 TGATGATACCACTTCACATATTACGAGACTGTATTTGTTGGGAAGAAAGGAAAGCGTCTCA 1562
Qy 1381 AGAAGAGTGGACCCCTTTTAAAGAGCAAAAAAGAAATTTTGAGAGAGAAAGCGGAAGCTT 1440
Db 1563 AGAAGATGGTCCCTTTTAAAGAGCAGAAAGAAATTTTGAGAGGAGAGACGAGCTT 1622
Qy 1441 TACAGAAGTGCATTCGATTCGGGTTGGAGAAAGGCGTTTGAAGAGAGCAGCCAG 1500
Db 1623 TACAGAAGCGGCTATTCGGCTGGGATTTGGAGAAAGGCAATTTGAAGAAAGAGCCAG 1682
Qy 1501 CTGGGTAAAGCAGAGTGTAAACATGACGAACCTTTGACACCCAGAACTCAGAAATGT 1560
Db 1683 TTGGTTTAAAGCAGAGTGTCTAAATATGACTACCTTTGACCAACCAAGAACTCAGAAATGT 1742
Qy 1561 GAAACTTTTTCAGTGCCCTCTCAGGAAGTCTGATCCAGACAATCTTATAGTCCACTCAGG 1620
Db 1743 GAAACTTTTTCAGTGCCCTCTCAGGAAGTCTGATTTGGGACATCTTATAGTCCACTCAG 1802
Qy 1621 GCCAGGCAAAAGAGCTACACAGTGTGGCTAATGGGTGCGAGCTTGACATCAAAACT 1680
Db 1803 GCAGCGCAAAAGAGCCCTCAGGTGTGTCTAATGGGTCTCCAGTTTGCATGTCTAAACT 1862
Qy 1681 GACTAAATCTCTCTGCTCAGCTCAGCTTCTACTTCAGACTTTGCCAGACACATTCATGTGT 1740
Db 1863 TACTAAATCTCTCTGCTCAGCTTCCACTTCCACTTCCAGACTTTTGCCAGACACGTTCTCGAT 1922
Qy 1741 GTCTGAACACAGTTCATCAGTGTGCTGAATAAATCTCTGAAGAAAGTAAACCAAGTGA 1800
Db 1923 ATCTGAACATAGTTCATCAATCAATGTAATATAAATCTGCTGAAGAAATTAACCAATCA 1982
Qy 1801 GGTTCGAAGAGAAAGCAGGATCAGAAAGTGGAGCGTGCAGTCGAGGCCAGCTCGCGGA 1860
Db 1983 GGTTCGAGGAGAACGTACAAATCAAAATGGAGTGTGGCGTCAAGACCTGGATCACAGA 2042
Qy 1861 GGGGTGCTACAGCGATGCTCTCGGCTTACGAGCGCTCAGGGGACCGAGATGACTT 1920
Db 2043 AGGTGCTATAGTGGATGCTCTCTGAGCTACAAATTTCTCATGTAGAAAAAGATGACTT 2102
Qy 1921 ACCTTAAATGTGCGGGCTGCAGTGTCTTCCAGATGTGGCTAGAGG-----A 1969
Db 2103 ACCTTAGACATGTGACGTGGAAATTTTTCATTAATGTGTTTCATCAAGTTTTCACATCTAA 2162
Qy 1970 GTTGAACAGGGGTGAGCAATAAGTCAGTCGTC-----TAACTTAAGATGCTCAGAGTTG 2024
Db 2163 GTTGAACAGGGGTGTGTATAAAGTCAGTTATCTCTAATACTTAAGATGGTCTGAGTTG 2222
Qy 2025 TTTGTTTGGACTTCGCTGCTCTCCCAAGAGCTGAAATGCT-AAGTACTTAAAGGA 2083
Db 2223 TTTGTTTGGACTTCCTGCTCTCTCCCAAGAGTTGAAATCTTAAATCTATTTTAAAGGA 2282
Qy 2084 TCGAAAGCTTTGG-----TTGTGTGTGTAGTAAACAGAGCCCTGGCTCTGTGACTGCAGGA 2139
Db 2283 TATAAAGCTTTGGATATGATTTTATAGTAACAGAAAGCATCTGGTTCTGTGAATAAAGGA 2342

Qy 2140 ATGCATGGC-GTTTGGATGGAAACAGAAAGCGCTGGAATGA--TTGCCTCGCCAGGTACCG 2196
Db 2343 ATGTATAGATGTTTGGATGGAAACAAAGCACTAGACTGAGTTTCTCTTATAGGTATTA 2402
Qy 2197 AGAAGAGCACTTTTGGAGGAGCTGCTCTGCTGTAACAACTTAAATATTCGTCCCAAGTGTGGTT 2256
Db 2403 AAAATAGCACTTTTGGAGAACTGATTAATTTGTAATGTTTAAATTTTGTCTCAAAATAGTT 2462
Qy 2257 GCATTGGAAGTGTAGCCCTTTACTTTGAATGTATCTGTAGATTTTAAACAAAGCAGGTTC 2316
Db 2463 GCATTGGAAGTGTAGCCCTTTACTTTGAATGTATCTGTAGATTTTAAACAAAGCAGGTTC 2522
Qy 2317 TATATTTATTTATGTTTGTGTTTGGGATTACCTCTCTTTCATATGTTTT----- 2367
Db 2523 TATATTTATTTATGTTTGTGTTG-GTTTGAATTTACCTCTTTCATATGTTTAAATAAAGT 2581
Qy 2368 -----GTGCTGTACATAAATATATACATGACTATGTAAAGAGCTTTTAAAGT 2415
Db 2582 GAAATTTATGATGTTTGTGTACATAGATACATGATTTATGTTAAAGAGCTTTTAAAGT 2641
Qy 2416 AAAAACTTTCACACCATGCTTGAGTATAGCATTTTCATGCCAATTAATAATGTTTTCAGTGGC 2475
Db 2642 AAAAGTTTCACA-CAACCATAGTATGATTTTCATGCC-AGTAAATTTTTCAGTGGT 2699
Qy 2476 ATGTTGTTTACAGAGG--TTAGGACCCTGCCACATGACAGTTAAAGACTTTTATTTTAAAG 2533
Db 2700 ATTCTGTTTACAGATGTTATAGGACCATTTGATGCAATACATTTAAGA-ATTCTCTTAAT 2758
Qy 2534 CCATCTGGGCAATAAAATTCAAAGCCCTTCAT-AAGCTGAGTTC---AGATAACTAGA 2589
Db 2759 ACATCTGGGCAATAAATAATTGAAGGTATTCATGAAGCTGAGTTCTTTAGATAATCAAC 2818
Qy 2590 ACTACTAAGTTTACATTTTTCAGATTTTAAAGCAATT-GTATTTTATTTATATATATGT-G 2647
Db 2819 ACTACTAATTTACATTTTTCAGATTTTTCAGATTTTATGACATTTAGATTTTATTTTGTATATGTAG 2878
Qy 2648 AATGTTATAATTTCTAAGAGGAATATTGATTATGGAGTAATGGGG 2692
Db 2879 AATATTATAATTTTAAAGGACTATTGATGATAGAAGAAATAGGG 2923

RESULT 9
AB023140
LOCUS Homo sapiens mRNA for KIAA0923 protein, partial cds. PRI 10-JAN-2004
DEFINITION
ACCESSION AB023140
VERSION AB023140.1 GI:4589477
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Hirose, M., Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O. Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
DNA Res. 6 (1), 63-70 (1999)
JOURNAL PUBMED 10231032
REFERENCE 2 (bases 1 to 5835)
AUTHORS Ohara, O., Nagase, T. and Kikuno, R.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-1999) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)
FEATURES
source
1..5835
/organism="Homo sapiens"
/mol_type="mRNA"

/db_xref="taxon:9606"
/clone="hh02882"
/sex="male"
/tissue_type="brain"
/clone_lib="pbluescriptii SK plus"
/dev_stage="adult"
1. 5835
/gene="KIAA0923"
CDS
4229..2109
/gene="KIAA0923"
/Note="Start codon is not identified"
/codon_start=1
/product="KIAA0923 protein"
/protein_id="BA76767.2"
/db_xref="GI:40788987"
/translation="TSSSGILALEIAMGDMVTDPGLSSSESKTISQVTSKTMSPSS
LYSQVLCSSIPLSKNVHSFFSAFCTEDNI EQSISYLDQELTTFGFSLSYBESKGET
KRENIIVAVLNCMVELLVQRNLLAENVETONLKGSDMDHILQSCYSLKEQLETS
RREMIGLQERORLOCKNRNLHLLKNEKDEVOKLONI IASRATQYNDHMKRREYN
KLKRLHLVNVKDKKIAMDILANYVGPADGKRGSMRTGKTEARNEDEMYKILLNDEY
YRQQLMENELKVLQOMKEMISLLSPQKKPRPRVDDSTGTVISDVEEDAGELS
RESMDLSCETVRQLTNSIRKQWRLKSHVEKLDNQVSKVHLEGFNDEDEVISQDHE
QETKLELEI OCKKEMIKTQOQLQOQLATAYDDDTLSLLDRCYLLBEKERLKEWEL
FKEQKNFERRRSFTAAIRLGLHKAFAFEERASWLKQOFLNNMTTFDHFQNSENVLKF
SAFGSSDWDNLI VHSROPKPHSVNSGSPVCMKLT KSLPASPSLTDPCOTRSCIS
EHSSINVLNITABEIKPNQVGECTNQKWSVASRPGSEGGYSCSLSYTNHVEKDD
LP"
ORIGIN
Query Match 63.1%; Score 1699.8; DB 5; Length 5835;
Best Local Similarity 80.6%; Pred. No. 0;
Matches 2213; Conservative 0; Mismatches 472; Indels 60; Gaps 17;
QY 1 CGTAGGAGGTGACAGGAGCTGTTGAAGCGTCGCAGCACCTGAGCGCGCTCCTCAGGTAT 60
DB 186 CATAGAAGAGTGACAGCAGTAGACTAAATGTTTAACTGCTCAACTAGTTCCTCAGGTAT 245
QY 61 CCTGCTCTGGAACCTGCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTAC 120
DB 246 CCTGGCTCTAGAGATTGCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCTGTCTTC 305
QY 121 AGAAACAAAAATCTCTCAATATACCTCAGAAA CAAAGATGTCTCCGTCACGTTTGTAT 180
DB 306 AGAAGCAAAACTATCTCTCAATATACCTCAGAA CAAAGATGTCTCCATCAAGTTTATA 365
QY 181 CTCACAGCAAGTTCTGTGCTCTTCTGATACCTTTATCCAAAACGTCGATGTGTTTCGG 240
DB 366 CTCACAGCAAGTGTCTGTTCTTCAATACCTTTATCGAAAATGTGCAAGTTTTCAG 425
QY 241 TGTCTTCTGCAGAGAGAA CATTGAA CAAAGTATTTCTTATCTTGATCAGGAGCTGAC 300
DB 426 TGCCTTCTGCACAGAGATATATATGACAGAGTATCTCATATCTTGATCAGGAATTGAC 485
QY 301 CACCTTCGGGTTTCTCTCTGTATGAAGATCCAAAAGTAAAGAGGCAAAAGAGAGAA TT 360
DB 486 TACTTTTGGTTTCTCTTCAATATATGAAGATCCAAAGGTAAAGAGACAAAGAGAGAGTT 545
QY 361 AAATATAGTCGTGTTCTGAACTGTATGAACGAGCTGCTGCTTCTGAGCGGAGAGAACT 420
DB 546 AAATATAGTAGCTGTACTAAATTTGATGAATGAGCTGCTTGTGCTTTCAGCGGAAGAACT 605
QY 421 GCTGCCCAGAGAGCGTGGAGACACAGAACTTGAAGCTGGCGAGTGACATGGACCACCT 480
DB 606 TCTAGCTCAGGANAATGTGGAGACACAGAAATTTGAGCTGGGAAGTATATGGACCATCT 665
QY 481 GCAGAGCTGCTACGCAAACTTAAAGGACAGTGTGAAAACGTCCAGCGGGAGATGATCGG 540
DB 666 ACAGAGCTGCTACTCAAAACTTAAAGGAACAACCTGGAAACCTCCAGAGGGGAAAATGATTGG 725
QY 541 GCTTCAAGAGAGACAGCGAGCTGCAGTGAAGNACAGGAGTTTGCATCAGCTCCTGAA 600
DB 726 GCTTCAGGAAAGAGACAGACAGTTACAAATGTGAAGNACAGGAATTTGCATCAGCTACTAAA 785

QY 601 GAATGAGAAAGATGAGGTACAAAAATTTACAAAATATCATAGCCAGCGGGCTACTCAGTA 660
DB 786 GAATGAGAAAGATGAGGTGCAAAAATTTACAAAATATCATTTGCAAGTGCAGCTACTCAGTA 845
QY 661 TAATCATGATGTGAAGAGGAGGCGTGAATATAATTAAGCTAAAGAGGCGCCTGCAATCA 720
DB 846 TAATCATGATGTGAAGAGAAAGAGCGTGAATATAATAA CTGAAGGAAACGCTCTACATCA 905
QY 721 GCTCGTTTATGAACNAGNAGGATAAAAACATAGCCATGATGTTTTAAATTTATGTGGGTG 780
DB 906 ACTTGTATTGAACNAGNAGGATTAAGAAAATATAGCTATGGACATTTTGAATTTATGTGCGGAG 965
QY 781 AGCTGATGGCAACCGAGGCTCATGAGGACATGACAAAACAGAAAGCCAGGAATGAAGATGA 840
DB 966 AGCTGATGGAAAAGAGGCTCTCTGGAGACTGGTAAAACTGAAGCCAGGAATGAAGATGA 1025
QY 841 GATGTACAAAATTTCTGTGAATGATTAATGATACCCGCAAGAGCAGATCTCTCTCTGTCTCC 900
DB 1026 AATGTA TAAAAATTTCTCTTGAATGATTAATGATTCGT CAGAAA CAAAATCCTTAATGGA AAA 1085
QY 901 CGCGGAGCTGAAGNAGTCTCTCAGCAGATGNAAGAGGAGATGATCTCTCTCTGTCTCC 960
DB 1086 TGCAGAACTTAAGAGGTTCTTCAACAAATGAAAAGGAAATGATTTCTCTCTTTCTCTCC 1145
QY 961 TCAGAAAGAAAGCCAGGGAAGAGCAGAGGACGGCACAGGCACTGTGCTATCTCCGA 1020
DB 1146 CCAAGAGNAGAACCTTAGANAGAGTAGATGATGATAGTACAGGAAC TGT---TATTTCCGA 1202
QY 1021 TATAGAAAGATGACTCTCTGGGAACTGAGCAGAGACAGCGTGTGGGGCTTTCTCTGTGACAC 1080
DB 1203 TGTGAAAGAGATGCCGGGAACTTAAGCAGAGAGAGTATGTGGGACCTTTCTCTGTGAAAC 1262
QY 1081 TGTGAGAGACAGCTGACAAACAGCATCAGNAAACAGTGAGAGATTTTGA AAAAGTCATGT 1140
DB 1263 TGTGAGAGAGCAGCTTACAAAACAGCATCAGAAAACAGTGAGAGATTTTGA AAAAGTCATGT 1322
QY 1141 AGAAAACTCGATAAACCAAGCTTCGAAGGTACACTCAGAGGGGCTTAAATGAGGAGGACGT 1200
DB 1323 AGAAAAGCTTGATAAACCAAGTTTCAAGGTACACTCGNAGGTTTTTAATGATGAAGATGT 1382
QY 1201 CATCTCAGCAAGACCATGAGCAAGAGACTGTAGAAAACCTGGAGCTGGAGATTGAGCGGTG 1260
DB 1383 AATCTCAGCAAGACCATGAA CAAAGAAA CTGAAAACTCGAGTTAGAAAATTCAGCAGTG 1442
QY 1261 TAAAGAGATCATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGCCACCACCGTGTGA 1320
DB 1443 TAAAGAAATGATTA AAACCTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGATATGA 1502
QY 1321 TGATGACACCACCTCACTGTTCGAGACTGTATTCTTCTCGAAGAAAAGGAAACGCTTAA 1380
DB 1503 TGATGATACCAC TTTCACTATTACGAGACTGTATTCTTGTGGNAGAAAAGGAAACGCTCAA 1562
QY 1381 AGAAGAGTGACCCCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAAGGCGAAAGCTT 1440
DB 1563 AGAAGAA TGGTCCCTTTTAAAGAGCAGAAAAGAAATTTTGAGAGGAGAGAGCAAGACTT 1622
QY 1441 TACAGAGCTGCCATTTCGATTTGGGTTGGAGAGAAAAGCGCTTTGAAGAAAGAGCGAGCCAG 1500
DB 1623 TACAGAAAGCCGCTATTTCGCTGGGATTGGAGAGAAAAGGCATTTGAGAGAGNAGAGCCAG 1682
QY 1501 CTGGGTAAGACAGCAGTTTTTAAACATGACGAACTTTTGACCACAGAACTCAGAAAAATGT 1560
DB 1683 TTGGTTAAAGCAGCAGTTTTCTAAAATATGACTACCTTTGACCACCAGAACTCAGAAAAATGT 1742
QY 1561 GAAACTTTTCAGTGCCTTTCTCAGGAAGTTCTGATCCAGACAAATCTTATAGTCCACTCAG 1620
DB 1743 GAAACTTTTCAGTGCCTTTCTCAGGAAGTTCTGATTTGGGACAAATCTTATAGTGCAC TCGAG 1802
QY 1621 GCCACGCAAAAGAGCTACACAGTGTGGCTTAATGGGGTGCAGCTTGCAGCATCAAAAAC T 1680
DB 1803 GCAGCCGCAAAAGAGCTCACAGTGTCTTAATGGGTCTCCAGTTTGCATGTCTTAACT 1862
QY 1681 GACTAAATCTCTTCTGSCCTCACCTTCTACTTTCAGACTTTTCGCCAGACACATTCATGTGT 1740

|||||
1863 TACTAAATCTCTCTGCTTCACTTCACTTCCAGACAGCTTCTCGAT 1922
QY
1741 GTCTGAACACAGTTCATCAGTGTGCTGAATATACTCTGAAGAAAGTAAACCAAGTGA 1800
Db
1923 ATCTGAACATAGTTCAATCAATGTACTGAATATACTGCTGAAGAAATTAACCAAAATCA 1982
QY
1801 GGTTCGACAGAGAACACGGATCAGAGTGGAGCGTGCAGTCGAGCGCCAGCTCCGGGA 1860
Db
1983 GGTTCGAGGAGAAATGTACAAATCAAAATGGAGTGGGCATCAAGACCTGGATACAGGA 2042
QY
1861 GGGGTGCTACACGGATGCTCTCGGCCCTTCAGGAGCGCTCACGGGACCGAGATGACTT 1920
Db
2043 AGTTTGTCTATAGTGTCTCTTGAGCTACAAATCTCATGTAGAAAGATGACTT 2102
QY
1921 ACCTTAAATGTGGGCTGCATGCTGTTCAGATGTCGCGTAGAGG-----A 1969
Db
2103 ACCTTAGACATGTGACTGGAATTTTTCATTAATGTGTTTCATCAAGTTTCACATCTAA 2162
QY
1970 GTTGACACAGGGGTAGCATAAAGTCAGTCGTC-----TAACCTTAAGATGCTCAGATG 2024
Db
2163 GTTGAACAGGGGTGTCTATAAGTCAGTTATCTCTAATACTTAAGATGGTCTGAGTTG 2222
QY
2025 TTTGTTTGGACTTCGCTGCTTCCCCCAAGAGCTGAAATGCT-AAGCTACTTAAAGGA 2083
Db
2223 TTTGTTTGGACTTCCTGCTTCCCCCAAGAGTTGAAATCTTAATCTATTTTAAAGGA 2282
QY
2084 TGCAAGCTTTGG-----TTGTGTGTAGTAAACAGAGCCCTGGCTCTGTGACTCGAGA 2139
Db
2283 TATAAAGCTTTGGATATGTATTTTGTAGTAAACAGAGCATCTGTTCTGTGAATAAGGA 2342
QY
2140 ATGCATGCG-GTTTGGATGGAACAGAGCGCTGGAATGA--TTGCCTCGCCAGGTACCG 2196
Db
2343 ATGTATAGATGTTTGGATGGAACAAAAGCACTAGACTGAGTTTCCTCTTATAGGTATTA 2402
QY
2197 AGAAGACACTTTTAGGACTGGTTCCTGTGTAACATTAAATATTCGTCGCCAGTGTGTT 2256
Db
2403 AAATAGACCTTTTAGGAACCTGATTAATGTAATGTTTAAATTTTGTCTCAATATAGTT 2462
QY
2257 GGCAATGGAAGTGTAGCCCTTTACTTGAATGTATCTGTAGATTTTAAACAGCAGGTTTC 2316
Db
2463 GGCAATGGAAGTTAGCCCTTTACTTGAATGTATCTGTAGATTTTAAACAGCGAGTTC 2522
QY
2317 TATATTTATATGTTTAGTGTGATTTTGGGATTACCTCTTTTCATATGTTT----- 2367
Db
2523 TATATTTATATGTTTAGTGTG-GTTTGAATTTACCTCTTTTCATATGTTTAAAAATAAGT 2581
QY
2368 -----GTGCTGTACATAAAATATACATGACTATGTTAAGAGCGCTTTAAGGTTT 2415
Db
2582 GAAATTTATGATGTTTGTACATAGATACATGATTTATGTTAAGAGCTTTAAGATTT 2641
QY
2416 AAAAATTTTACACCATGCTTGAAGTATAGCATTTTCAATGCAATTTAAATGTTTTCAGTGGC 2475
Db
2642 AAAAGTTTTCACA-CAACCATAGTATAGTATTTTCATGCC-AGTAAATTTTTTTAGTGT 2699
QY
2476 ATGTTGTTTACAGG--TTAGGACCACTGCCACATGACAGTTAAGACTTTATTTTAAAG 2533
Db
2700 ATTCTGTTTACAGATGTATAGGACCAATTGATGCAATTACATTTAAGA-ATTCTCTTTAAT 2758
QY
2534 CCATCTGGGCAATAAAATTCAAAGCCCTTCAT--AAGCTGAGTTC---AGATAACTAGA 2589
Db
2759 ACATCTGGGCAATAAATTTGAAGGTATTCATGAAGCTGAGTTCTTTAGATTAATCAAC 2818
QY
2590 ACTACTAAGTTTACATTTTGGAGATTTTAAAGCAATT-GTATTTTATTTTATATATGT-G 2647
Db
2819 ACTACTAATACATTTTGGAGATTTTATGACATATTAGATTTTATTTTGTATATGTAG 2878
QY
2648 AATGTTTAAATTTCTAAGAGGAATTTGATTTATGAGGATTAATGGGG 2692
Db
2879 AATATTATAATTTTAAAGGAGCTATTGATGATGAAGAATAGGG 2923

RESULT 10

AY651262
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .3073
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="p22.3"
<1. .3073
/gene="SSX2IP"
/gene="SSX2IP"
/note="similar to synovial sarcoma X breakpoint 2
interacting protein; contains possible frameshift"
ORIGIN
Query Match 56.2%; Score 1512.2; DB 5; Length 3073;
Best Local Similarity 80.4%; Pred. No. 0;
Matches 2023; Conservative 0; Mismatches 428; Indels 64; Gaps 19;
Qy 231 GTGTTTTCGGTGTCTTCTGCACAGGAGAACATTGAACAAAGTATTTTCTTATCTTGATC 290
Db 1 GTTTTTTCAGTGCCTTCTGCACAGAGATAATATTGAACAGAGTATCTCATCTTGATC 60
Qy 291 AGGAGCTGACCACTTCGGGTTTCTTCTTGTATGAAGATCCAAAAGTAAAGAGCAA 350
Db 61 AGGAATTGACTACTTTTGGTTTTCTTCAATTATATGAAGATCCAAAGGTAAGAGACA 120
Qy 351 AGAGAGAAATAATATAGTCGCTTCTGAACCTGTATGAAGAGCTGCTCGTCTTCAGC 410
Db 121 AGAGAGATTAAATATAGTAGTACTATAATTGTATGAATGAGCTGCTGTGCTTCAGC 180
Qy 411 GGAAGAACCTCTGCTGGCCAGGAGCGTGGAGACACAGAACTTGAAGCTGGGCACTGACA 470
Db 181 GGAAGAACCTTCTAGCTCAGGAAATCTGGAGACACAGAACTTGAAGCTGGGAGTGATA 240
Qy 471 TGGACCACTGCAGAGCTGTACGCCAAACTTAAGAGAGAGTTGGAAACGTCCAGCGGG 530
Db 241 TGGACCACTGCAGAGCTGTACTCAAAACTTAAAGGAACAACTGGAACCTCCAGGAGG 300
Qy 531 AGATGATCGGGCTTCAAGAGAGAGACAGGAGCTGCAGTGCAGAACACAGAGATTGATC 590
Db 301 AATGATTGGCTTTCAGAAAGAGACAGACAGTTTACAATGTAAAGAACAGGAATTGATC 360
Qy 591 AGCTCTTGAAGAATGAGAAAGATGAGGTACAAAAATTTACAAAATATCATAGCCAGCGGG 650
Db 361 AGCTACTAAGAAATGAGAAAGATGAGGTGCANAAATTTACAAAATATCATTCAGAGTCAG 420
Qy 651 CTACTCAGTAAATCATGTGTGAAGAGGAGCGTGAATATAATAAGCTAAAGAGC 710

|||||
421 CTACTCAGTATATCATGNTATGATGAAGAGAAAGAGCGTGATATATATAAATCTGAAGGAC 480
QY
711 GCCTGCATCAGCTCGTTATGAAACAAAGAGGATAAAAAACATAGCATGCCATGTTTAAATT 770
Db
481 GTCTACATCAACTGTTTATGAACAAGAAAGATAAGAAAAATAGCTATGGACATTTTGAATT 540
QY
771 ATGTGGGTGAGCTGTAGTGGCAACGAGGCTCATGGAGGACTGACAAAACAGAACGCCAGGA 830
Db
541 ATGTGGGAGAGCTGTAGTGGAAAAAGAGGCTCTGGAGGACTGGTAAAACTGGAAGCCAGGA 600
QY
831 ATGAAGATGAGATGTACAAAAATCTGTTTGAATGATTTAGTACCGCAGAGAGCAGATCC 890
Db
601 ATGAAGATGAATGTATAAAAATCTCTTGAATGATTTAGTATATCTGCAGAAAACAAATCC 660
QY
891 TGATGGAGAACCGGAGCTGAAGAGTCTCTCCAGCAGATGAAGAAGAGATGATCTCTC 950
Db
661 TAATGGAAAAATGCAGAACTTAAGAAGGTTCTTCAACAAATGAAAAAGAAATGATTTCTC 720
QY
951 TCCTGTCTCCTCAGAAAGAGGCCAGGGAAGAGCAGAGGACGGCAGACGCACTCTTG 1010
Db
721 TTCCTTCTCCCAAAAGAAAGAAACCTAGAGAAAGAGTAGATGATAGTACAGGAACCTGT-- 778
QY
1011 CTATCTCCGATATAGAAGATGACTCTGGGGAACCTGAGCAGAGACAGCGTGTGGGGCCCTTT 1070
Db
779 -TATTTCCGATGTTGAAGAAGATGCCGGGAACCTAAGCAGAGAGATGATGTGGACCTTT 837
QY
1071 CCTGTGACATGTGAGAGAGCAGCTGACAAACAGATCAGGAAACAGTGGAGAAATTTGA 1130
Db
838 CTTGTGAACTGTGAGAGAGCAGCTTACAAACAGCATCAGAAAAACAGTGGAGAAATTTGA 897
QY
1131 AAAGTCATGTAGAAAACTTCGATACCAAGCTTCGAAAGGTACACTCAGAGGCGCTTAATG 1190
Db
898 AAAGTCATGTAGAAAAAGCTTCGATACCAAGTTTCAAAGGTACACTCAGAGGCTTTAATG 957
QY
1191 AGGAGGAGCTCATCTCAGCAAGAACCATGAGCAAGAGACTGAGAAACCTGAGCTGGAGA 1250
Db
958 ATGAAGATGTAATCTCAGCAGACAGACCATGAACAGAACTGAAAACTCGATTAGAAA 1017
QY
1251 TTGAGCGGTGTAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCA 1310
Db
1018 TTCAGCAGTGAAGAAATGATTAATACTCAGCAACAGCTTTTA--CAGCAGCTCGCTA 1074
QY
1311 CCAGTGTGATGATGACACCACTCAGCTGTTGCGAGACTGTTACTTGTGCGAAGAAAGG 1370
Db
1075 CTGCATATGATGATGATACCACTTCACATATACGAGACTGTTATTTGTTGGAGAAAGG 1134
QY
1371 AACGCTTTAAAGAGAGTGGACCTTTTAAAGAGCAAAAAAGAAATTTTGAGAGAGAAA 1430
Db
1135 AACGTCTCAAGAGAAATGGTCCCTTTTAAAGAGCAGAAAAAGAAATTTTGAGAGGAGA 1194
QY
1431 GCGAAGCTTTACAGAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGTTTGAAGAG 1490
Db
1195 GACGAAGCTTTACAGAAAGCGCTATTCGCTGGGATTGGAGAGAAAGCAATTTGAAGAG 1254
QY
1491 AGCGAGCCAGCTGGTAAAGCAGCAGTTTAAACATGACCACTTTTGACCAACAGCAACT 1550
Db
1255 AAAGAGCCAGTTGGTTAAAGCAGCAGTTTCAAATATGACFACCTTTTGACCAACAGCAACT 1314
QY
1551 CAGAAATGTGAAAATCTTTTCAGTGCCTTCTCAGGAAGTTTCTGATCCAGCAATCTTATAG 1610
Db
1315 CAGAAATGTGAAAATCTTTTCAGTGCCTTCTCAGGAAGTTTCTGATTTGGGACAACTTATAG 1374
QY
1611 TCCACTACGGCCAGCGGAAAAGAGCTACACAGTGTGGCTTAATGGGTCGACGCTTGCA 1670
Db
1375 TGCATCGAGGCGCGGAAAAGAGCCTCACAGTGTGTCTAATGGGTCTCCAGTTTGCA 1434
QY
1671 CATCAAACTGACATAAATCTCTCTGCTGCTCACCTTCTACTTCTGACACTTTCGCGACAGAC 1730
Db
1435 TGCTTAAACTTACTAAATCTCTCTCTGCTTCCCTTCCACTTCAGACTTTTGGCCAGACAC 1494
QY
1731 ATTCAATGTGTCTGAACACACAGTTCCATCAGTGTGCTGAATATAACTCTCTGAAAGAAAGTA 1790
|||||

Db
1495 GTTCTGCATATCTGAACATAGTTCAATCAATCAATGACTGAATAATAACTGCTGAAGAAATTA 1554
QY
1791 AACCAAGTGAAGTTGCAAGAGAAAGCACGGATCAGAGTGGAGCGTGCAGTCGAGGCGCCA 1850
Db
1555 AACCAAAATCAGGTTGGAGGAGAACGTACAAATCAAAATCAAAATGGAGTGTGGCGTCAAGACCTG 1614
QY
1851 GCTCGGGGAGGGGTGCTACAGCGGATGCTCTCGGCCCTTCAGAGCGCTCACGGGAGCC 1910
Db
1615 GATCAGAGG-AGTGCTATAGTGGATGCTCTTGGAGTACACAAATCTCATGTAGAAA 1673
QY
1911 GAGATGACTTACCTTAAATGTGCGGGTGCAGTGTCTCCAGATGTGCGGTAGAGG-- 1968
Db
1674 AAGATGACTTACCTTAGACATGTGGACTGGAAATTTTTTTCATTAATGTGTTCAAGTT 1733
QY
1969 -----AGTTGACACAGGGGTAGCATAAAGTCAAGTCAAGTCTGCTC-----TAACTTAAAGATG 2014
Db
1734 TCACATCTAAGTTTGAACAGGGGTGTGTATAAAGTCAAGTTATCTCTAATAACTTAAAGATG 1793
QY
2015 CTGAGAGTCTGTTGTTTGGACTTTCGCTGCTCTCCCAAGAGCTGAAATGCT--AAGCTA 2073
Db
1794 GTCTGAGTGTGTTTGGACTTTCCTGCTCTTCCCAAGAGTGAATCTTTAAATCTA 1853
QY
2074 CTTTAAAGAGATGCAAGCTTTTGG-----TTGTGTGTGTAGTAAACAGAGCCCTGCTCTGT 2129
Db
1854 TTTTAAAGAGATATAAAAGCTTTGGATATGATTTTGTAGTAAACAGAGCATCTGGTTCTGT 1913
QY
2130 GACTGCAGGAATGCAT--GGCGTTTGGATGGAACAGAGCGCTGGAGTGA--TTGCGCTCG 2186
Db
1914 GAATAAGGAATGTATAGATGTTTGGATGGAACAAAGCAGCTAGCTGAGTTTCTCTT 1973
QY
2187 CAGGTCACCGAGAGAGCAGCTTTTAGGACTGGTTCCTGTAAACATTTAAATATTTCTGCTCC 2246
Db
1974 ATAGTATTTAAATAAGTCTTTTAGGAACTGATTTATTTGTAATTTTAAATTTTGTCTC 2033
QY
2247 AAGTGTGTTGGCATTGGAAGTGTAGCTTTTACTTGAATGTATCTGATAGATTTTAAACA 2306
Db
2034 AAATATAGTTGGCATTGGAAGTTTAGCTTTTACTTGAATGTATCTGATAGATTTTAAACA 2093
QY
2307 AACGAGTCTCATATTTATTTATGTTAGTGTGATTTTGGGATTTACCTTTTTCATATGTTT 2366
Db
2094 AAGCGAGTCTCATATTTATTTATGTTTAGTGTG--GTTTGAATTTTACCTCTTTTCATATGTTT 2152
QY
2367 T-----GTGCTGTACATAAATATACATGACTATGTTTAAAGAGGC 2405
Db
2153 TAAATAAGTGAAATTTATGATGTTTGTGTACATAGATACATGATGTTTAAAGAGGC 2212
QY
2406 TTTAAGGTTTAAACCTTTCACACATGCTTGTAGTATAGCAATTTTCATGCCAATTTAAATGT 2465
Db
2213 TTTAAGATTTTAAAGTTTTCACA--CAACCATAAAGTATAGTATTTTCATGCC--AGTAAAAATTT 2270
QY
2466 TTTTCAGTGGCATGGTGTTTTACAGAG--TTAGGACCACTGCCACATGACAGTAAAGACTT 2523
Db
2271 TTTTAGTGGTATTTCTGTTTACAGATGATTTAGGACCAATTTGATGCAATTTACATTTAAGA-AT 2329
QY
2524 TATTTTAAAGCCATCTGGGCAATAAAATTTCAAAGCCCTTTCAT--AAGCTGAGTTC--A 2579
Db
2330 TCTCTTTAATACATCTGGGCAATAAATATTGAAAGGTATTTCCATGAAGCTGAGTCTCTTA 2389
QY
2580 GATACTAGAACCTACTAACCTTACATTTTGTAGATTTTAAAGCAATTTGATTTTATTTT 2638
Db
2390 GATAATCAACACTACTAACATTTTGTAGATTTTGTAGATTTTATGACATTTAGATTTTATTTT 2449
QY
2639 ATATATGT--CAATGTTTATAAATTTCTTAAGAGGAATTTGATTTATGAGTAATGGGG 2692
Db
2450 GTATATGTAGAAATATATAAATTTTAAAGGACTATTTGATGATAGAGAAATAGGG 2504
|||||

RESULT 11
AY367055
LOCUS
DEFINITION Homo sapiens synovial sarcoma, X breakpoint 2 interacting protein (SSX2IP) mRNA, complete cds.
ACCESSION AY367055

VERSION	AY367055.1	GI:34451652	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1. (bases 1 to 2379)		
AUTHORS	Lu, L., Huang, X. Y., Xu, M., Yin, L. L., Li, J. M., Zhou, Z. M. and Sha, J. H.		
TITLE	Cloning a new transcript of X breakpoint 2 interacting protein(SSX2IP) in testis		
JOURNAL	Unpublished		
REFERENCE	2. (bases 1 to 2379)		
AUTHORS	Sha, J. H., Zhou, Z. M. and Li, J. M.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-AUG-2003) Key Lab of Reproductive Medicine, Nanjing Medical University, 140 Han Zhong Road, Nanjing, Jiangsu 210029, China		
FEATURES	Location/Qualifiers		
source	1. .2379		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/chromosome="11"		
	/tissue_type="testis"		
gene	1. .2379		
	/gene="SSX2IP"		
CDS	190. .2034		
	/gene="SSX2IP"		
	/codon_start=1		
	/product="synovial sarcoma, X breakpoint 2 interacting protein"		
	/protein_id="AAQ72373.1"		
	/db_xref="GI:34451653"		
	/translation="MGDAWTVTDPLGSSSEKTIISQYTSETKMSPSLYSOVLCSSIP LSKVHSFSACTEDNI EQSILYDLQELTFPGPSLYEESKSKHETKRELNIYAVLNC MNEILLVQRKULQAQVETQNLKUGSDMDHLQSCYSLKEQLETSREMIQOERDR QOCCRNRLHOLLNKEDEVKQLNIIASRATQYNHDMKREYNNKLKERLHQLVNN KDKQIAMILNAGRADCKGRWRTGTEARNEDMYKILLNDEYRQKIIIMENAE LKKVLQOMKEMISLLSPKKPRVRVDDSTGTVISDVEEDAGELSNMWDLSCETV RQOLNSIRKQWRILKSHVEKLDNOVSKVHLEGENDVDVSRDPHEQETKLEILQQ CEMIKTQOQLLOQLATAYDDTTSLLRDCYLLEERLKEBSLFRQKQKPERER RFTBAAIKGLERKAFBEERASWLKQOFLNMFTFDHONSENVKLFSSHSSDWNLI IVHSRQPKHVSNGSPVCMVSKLTKLSPASPTSDFCQTRSCISFSHSINVLNITA EBIKPNQVGEVCTNOKWSVASRPSQEGCYSGCSLSYTNSHVEKDLP"		
ORIGIN			
Query Match	55.4%;	Score 1492.6;	DB 5; Length 2379;
Best Local Similarity	81.4%;	Pred. No. 0;	
Matches 1823;	Conservative 0;	Mismatches 389;	Indels 27; Gaps 7;
QY	1	CGTAGGAGAGTGCAGGAGCTGCTGTAAGCGTCGCGAGCACCTGAGCCGCTCCTCAGGTAT	60
DB	111	CATAGAAGAGTGCAGCAGCTAGACTAAATGTTTAACTGCTGAACTAGTTCCTCAGGTAT	170
QY	61	CCTGGCTCTGGAACTTGCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTAC	120
DB	171	CCTGGCTCTAGAGATTGCTATGGGAGATTGGATGACTGTACAGATCCAGTCTGTCTTC	230
QY	121	AGAAAAAATAATCTCTCTCAATATATACCTCAGAAAAAAGATGTCTCGTCCAGTTTGTGA	180
DB	231	AGAAAGCAAAACTATCTCTCAATATACCTCAGAAAAAAGATGTCTCCATCAAGTTTATA	290
QY	181	CTCCAGCAGCAAGTCTGTGCTCTTACGTACCTTTATCCAAAAACGTGGATGGTGTTCGG	240
DB	291	CTCAGCAGCAAGTCTGATGTTCTTCAATACCTTTATCGAAAAAATGTGCACAGTTCCTCAG	350
QY	241	TGTTCTTCTGCAGCAGGAGAGAAATATGAACAAAGTATTTTCTTATCTTGATCAGGAGCTGAC	300
DB	351	TGCTTTCTGCACAGAGATAATTTGAACAGAGTATCTCATATCTTGATCAGGAATTGAC	410
QY	301	CACCTTCGGGTTTCTTCTTGATATGAAGAAATCCAAAAAGTAAAGGCAAGAGAGAAATT	360

```
Qy 1441 TACAGAAGCTCCATTGATTTGGGTTGGAGAGAAAGCGTTTGAAGAGCGAGCCAG 1500
Db 1548 TACAGAAGCGCCCTATTTCGCTGGGATTGGAGAGAAAGCGCATTTGAAGAAGAAAGAGCCAG 1607
Qy 1501 CTGGGTAAGAGCAGAGCTTTTAAACATGACGAACTTTGACCAACAGAACTCAGAAAATGT 1560
Db 1608 TTGGTTAAGAGCAGAGCTTTTAAATATGACTACTTTGACCAACAGAACTCAGAAAATGT 1667
Qy 1561 GAAACTTTTTCAGTCCCTTCTCAGGAAGTTTGATCCAGACAAATCTTATAGTCCACTCAAG 1620
Db 1668 GAAACTTTTTCAGTCCCTTCTCAGGAAGTTTGATCCAGACAAATCTTATAGTCCACTCGAG 1727
Qy 1621 GCCACGGCAAAAGAGCTACAGATGGCTTAATGGGGTGCAGCTTGCACATCAAAACT 1680
Db 1728 GCAGCCGCAAAAGAGCGCTCAGGTGTCTTAATGGGTCTCCAGCTTTGCAATGTCTAAACT 1787
Qy 1681 GACTAAATCTCTCTCGCTCCTCACCTTCTACTTTCAGACTTTGCCAGACACATTCATGTGT 1740
Db 1788 TACTAAATCTCTCTCGCTCCTCACCTTCCACTTCAGACTTTGCCAGACACGTTCTGCTCAT 1847
Qy 1741 GTCTGAACACAGTTCCATCAGTGTGCTGAATATATACTCTGAAGAAAGTAAACCAAGTGA 1800
Db 1848 ATCTGAACATAGTTCAATCAATGACTGCTGAATATAACTGCTGAAGAAATTAACCAAAATCA 1907
Qy 1801 GGTTCGAAGAAAGACGAGATCAGAGTGCAGTGCAGTGCAGGCGCCAGCTCCCGGA 1860
Db 1908 GGTTCGAGTGAATGTACAAATCAAAATGAGTGTGGCATCAAGACCTGGATCACAGGA 1967
Qy 1861 GGGGTGCTACAGCGGATGCTCTCGGCTTTCAGAGCGCTCACGGGACCGAGATGACTT 1920
Db 1968 AGGTGCTATAGTGGATGCTCTTGGAGTACACAAATTCATGATAGAAAAGATGACTT 2027
Qy 1921 ACCTTAAATGTCGGGCTGCAGTGTCTTCCAGATGTCGCTAGAGG-----A 1969
Db 2028 ACCTTAGACATGTGGACTGGAATTTTTCATTAATGTTGTTTCATCAAGTTTCACATCTAA 2087
Qy 1970 GTTACACAGGGTGTAGCATAAAGTCAGTCGTC-----TACTTAAGATGTCAGAGTTG 2024
Db 2088 GTTGAACAGGGTGTGTCAAGTCAAGTATCTTCTAATAACTTAAAGATGCTGTGAGTTG 2147
Qy 2025 TTGTTTGGACTTCGCTCTCTTCCCCAAAGAGCTGAAATGCT-AAGCTACTTAAAGGA 2083
Db 2148 TTGTTTGGACTTCCTGCTCTTCCCCAAAGAGTGAATCTTAAATCTTATTAAGGA 2207
Qy 2084 TGCAAAAGCTTTGG-----TTGTGTGTAGTATACAGAAGCCCTGCTGTGACTGTCAGGA 2139
Db 2208 TATAAAGCTTTGGATATGATTTTGTAGTACAGAAGCATCTGCTTCTGTGAATAAAGGA 2267
Qy 2140 ATGCAT-GGCGTTTGGATGGAACAGAGCGCTGGAATGA--TTGCCTCGCCAGGTACCG 2196
Db 2268 ATGTATAGATGTTGGATGGAACAAAGCACTAGACTGAGTTTCTCTTATAGGTATTA 2327
Qy 2197 AGAAGAGCACTTTTAGGGA 2215
Db 2328 AANATAGCACTTTTAGGAA 2346

RESULT 12
LOCUS BC064389
DEFINITION Homo sapiens synovial sarcoma, X breakpoint 2 interacting protein,
mRNA (CDNA clone MGC:75026 IMAGE:5745142), complete cds.
ACCESSION BC064389
VERSION BC064389.1 GI:39963669
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 2540)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
```

```
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zebberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Haieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Udwin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalski,U., Smalusi,D.E.,
Schnerk,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2540)
Director MGC Project.
Direct Submission
Submitted (15-DEC-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-x@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc\_mgc@hgrl.nih.gov
Akhtar,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Mapkari,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Turgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAK Plate: 139 Row: k Column: 13
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 41281570.
FEATURES
Location/Qualifiers
1..2540
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:75026 IMAGE:5745142"
/tissue_type="Brain, adult medulla"
/clone_lib="NIH MGC_119"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1..2540
/genes="SSX2IP"
/notes="synonyms: KIAA0923, ADIP"
/db_xref="GeneID:117178"
/db_xref="MIM:608690"
241..1920
/genes="SSX2IP"
/codon_start=1
/product="SSX2IP protein"
/protein_id="AAH64389.1"
gene
CDS
```

		/db_xref="GI:39963670"	
		/db_xref="GeneID:117178"	
		/db_xref="MIM:608690"	
		/translation="MGDMVTDPGLSSSESKTSIQVTSRKMSPSLSXQVLCSSIP LSKNVHSFFSAFTCEDNIQSISYLDQELTTFGFPSPLYEESKGETKRELNIVAVLNC MNELLVLRQLNLLAQNENKDESVOKLONLIIASRATQYNHDKMKEREYKULKERLHOLVDR OLOCKNRNLHLLAQNENKDESVOKLONLIIASRATQYNHDKMKEREYKULKERLHOLVDR KDKKIAMDILNIVGRADGKRGWRTGTEARNEDEMYKILLNDYEYKQQLIMENAE LKVLQOMKEMISLSPQKKPRERVDSTGTVISDVEEDAGELSRSMWDLSCETV RQOLNIRKQWR11LKSHVEKLDNQVSKVHLEGFNDEDEVISQDHEQTEKLELEIQ CHEMTQOQLQALATAYDDDTTSLLRDCYLLBEKRLKEEWSLFKQKKNFERRR SFTEAAIRLGLSRKAFEEERASWLKQOFLNMTTFDHOENSENVKLFSAFGSSDMDNLI VHSRQPKPHSVNSGSPVCMKSLTKSLPASPSSTDFTQTRSCISEHRAKN"	
ORIGIN			
Query Match		47.9%; Score 1290.6; DB 5; Length 2540;	
Best Local Similarity		83.1%; Pred. No. 0;	
Matches 1496; Conservative		0; Mismatches 299; Indels 6; Gaps 2;	
QY	1	CCTAGGAGTGCACAGGAGCTGTGTAAAGCGTGGCAGCACTGAGCGGCCTCTCTCAGGTAT	60
DB	162	CATAGAAGAGTGACAGCAGCTAGACTAAATGTTAACTGCTGAACCTAGTGTCTCAGGTAT	221
QY	61	CCTGGCTCTGGAACCTTGCTATGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTAC	120
DB	222	CCTGGCTCTAGAGATTGCTATGGGAGATTGGATGACTGTTCAGATCCAGGTCTGTCTTC	281
QY	121	AGAAAACAAAAACTCTCTCAATATACCTCAGAAAAAAGATGTCTCGCTCCAGTTTGTGA	180
DB	282	AGAAAGCAAAAATCTCTCTCAATATACCTCAGAAAAAAGATGTCTCCATCAAGTTTATA	341
QY	181	CTCCAGCAAGTCTGTGCTCTCTAGTACCTTTATCCAAAAACGTGCGATGGTGTTCGG	240
DB	342	CTCAGCAAGTGTGTATGTTCTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTCAG	401
QY	241	TGCTCTCTGCAAGGAGAAACATTTGAACAAAGTATTTCTATCTTGATCAGAGCTGAC	300
DB	402	TGCTCTCTGCAAGGAGAAATATTTGAACAGAGTATCTCATATCTTGATCAGAAATTGAC	461
QY	301	CACCTTCGGGTTTCTCTTGATGAAGAATCCAAAAAGTAAAGAGGCAAAAGAGAGAATT	360
DB	462	TACTTTTGGTTTCTCTTATATGAAGAATCCAAAGTAAAGAGACAAAGAGAGATT	521
QY	361	AAATATAGTCGTCTGTGAACCTGTATGAAACGAGTCTGTGCTTTCAGCGGAAGAACT	420
DB	522	AAATATAGTCGTGTACTAAATTTGTATGAATGAGCTGTGTGCTTTCAGCGGAAGAACT	581
QY	421	GCTGCCAGGAGAGCGTGGACACAGAACTTTGAAGCTGGGACGTGACATGGACCACT	480
DB	582	TCTAGCTCAGGAAATGTGGAGACACAGAAATTTGAAGCTGGGAAGTATGGACCATCT	641
QY	481	GCAGAGCTGTACTACGCAAACTTTAAGGAGCAGTTGGAAAGCTCCAGCGGGAGATGATCGG	540
DB	642	ACAGAGCTGTACTCAAAACTTTAAGGAACAACCTGGAAACCTCCAGGAGGGAATGATTTGG	701
QY	541	GCTTCAAGAGAGACAGCGAGCTGCGATGCAAGAACAGGAGTTTGCATCAGCTCCTGAA	600
DB	702	GCTTCAAGAAAGAGACAGACAGTTTCAATGTGAAGAACAGGAATTTGCATCAGCTACTAAA	761
QY	601	GAATCAGAAAGATGAGGTACAAAATTTACAAAATATCAATATTAAGCTTAAAGGAGCGCTGCATCA	660
DB	762	GAATCAGAAAGATGAGGTGCAAAAATTTACAAAATATCAATATTAAGCTTAAAGGAGCGCTGCATCA	821
QY	661	TAATCATGATGTGAAGAGAGGAGCGTGAATATTAATTAAGCTTAAAGGAGCGCTGCATCA	720
DB	822	TAATCATGATGTGAAGAGAGAGCGGTGAATATTAATTAAGCTTAAAGGAGCGCTGCATCA	881
QY	721	GCTGCTTATGAACAAGAGAGTAAAAACATAGCCATGGATGTTTAAATATATGTGGGTG	780
DB	882	ACTTGTATGAACAAGAGAGTAAAAACATAGCCATGGATGTTTAAATATATGTGGGTG	941
QY	781	AGCTGATGCAACAGGCTCATGAGGACTGACAAAACAGAAAGCCAGGAATGAAGATGA	840

RESULT 13
BD156617

DB	942	AGCTGATGGAAGAAAGAGGCTCTCGAGGAGCTGGTAAAACTGAAGCCAGGAATGAAGATGA	1001
QY	841	GATGTACAAAATTTCTGTTGAATGATTATGAGTACCGCCAGAAAGCAGATCCTGATGGAGAA	900
DB	1002	AATGTATAAAATTTCTCTTGAATGATTATGAAATATCGTCAGAAACAAATCTTAATGGAAA	1061
QY	901	CGCGAGCTGAAGAAGGCTCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTGTCTCTCC	960
DB	1062	TGCAGAACTTAAGAAGGTTCTTCAACAAATGAAAAAGGAAATGATTTCTCTCTCTCTCTCC	1121
QY	961	TCAGAAAGAAAGCCAGCGGAAAGAGCAGAGGACCGCAACAGGCACTGTTGCTATCTCCCA	1020
DB	1122	CCAAAGAAAGAAACCTAGAGAAAGAGTAGATAGTACAGGAACCTGT---TATTTCGA	1178
QY	1021	TATAGAAGATGACTCTCGGGAACTCGAGCAGAGACAGCGTGTGGGGCCCTTTCTCTGTGACAC	1080
DB	1179	TGTTGAAGAGATGCGCGGGAACCTAAGCAGAGAGATATGTGGGACCTTTCTCTGTGAAAC	1238
QY	1081	TGTGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAGATCATGT	1140
DB	1239	TGTGAGAGCAGCTTACAAAACAGCATCAGAAACAGTGGAGAAATTTTGAAGATCATGT	1298
QY	1141	AGAAAACTCGATTAACCAAGCTTCGAAGGTACACTCAGAGGSCCTTAATGAGGAGACCT	1200
DB	1299	AGAAAGCTTGATAACCAAGTTTCAAGGTACACTCGGAAGGTTTAAATGATGAAGATGT	1358
QY	1201	CATCTCAGCAAGACCATGAGCAGAGAGATGAGAAACCTGAGAGCTGGAGATTGAGCGGTG	1260
DB	1359	AATCTCAGCAAGACCATGAAACAGAACTGAAAACTCGAGTTAGAAATTTACAGAGTG	1418
QY	1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACACGTGTGA	1320
DB	1419	TAAAGAAATGATTAATAACTCAGCAACAGCTTTTA---CAGCAGCTCGCTACTGCAATGA	1475
QY	1321	TGATGACACCACTCAGCTGTCGAGAGCTGTTACTTCTGCTGGAAGAAAAAGGAACGCTTAA	1380
DB	1476	TGATGATACCACTTCACTATTTACGAGACTGTTATTTGTTGGAAGAAAAAGGAACGCTCAA	1535
QY	1381	AGAAAGTGGACCCCTTTTAAAGAGCAAAAAAGAAATTTGAGAGAGAAAGCGCAAGCTT	1440
DB	1536	AGAAAGTGGTCCCTTTTAAAGAGCAGAAAAAGAAATTTGAGAGGAGAGACGAAGCTT	1595
QY	1441	TACAGAGCTGCCATTCGATTTGGGGTTGGAGAGAAAGCGGTTTGAAGAAGAGCGAGCCAG	1500
DB	1596	TACAGAGCGCTATTCGCTGGGATTGGAGAGAAAGGCAATTTGAAGAAGAAAGAGCCAG	1655
QY	1501	CTGGTAAAGCAGCAGTTTTTAAACATGACCGAACTTTGACCAACCAAGAACTCAGAAAAATGT	1560
DB	1656	TTGGTTAAAGCAGCAGTTTTCTAAATATGACTACCTTTGACCAACCAAGAACTCAGAAAAATGT	1715
QY	1561	GAACCTTTTTCAGTGCCTTCTCAGGAAGTTCTGATCCAGACAATCTTATAGTCCACTCAGC	1620
DB	1716	GAACCTTTTTCAGTGCCTTCTCAGGAAGTTCTGATTTGGGACAATCTTATAGTCACTCGAG	1775
QY	1621	GCCACGGCAAAAGAGCTACACAGTGTGGCTTAATGGGGTGCAGCTTGCAATCAAAACT	1680
DB	1776	GCAGCGCAAAAGAGCGCTCAACAGTGTGCTTAATGGGTCTCCAGTTTGCATGTCTAAACT	1835
QY	1681	GACTAAATCTCTCTCGCTCAGCTTCTTACCTTCTGAGCTTTCCGAGACACATTCATGTGT	1740
DB	1836	TACTAAATCTCTCTCGCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT	1895
QY	1741	GTCTGAACACAGTTCCTCAGTGTGCTGAATATAAATCTCCTGAAGAAAGTAAACCAAGTGA	1800
DB	1896	ATCTGAACATAGGGCAAGAACTCAGAAAAACGCTGTTCTCTGTTGAGTGAAGAAAGTTTGT	1955
QY	1801	G 1801	
DB	1956	G 1956	

LOCUS BD156617 2716 bp DNA linear PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD156617
VERSION BD156617.1 Gi:27862375
KEYWORDS JP 2002191363-A/11460.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 2716)
Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Suiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002191363-A 11460 09-JUL-2002;
JELIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/11460
PD 09-JUL-2002
PF 28-JUL-2000 JP 2002080990
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU
SAITO,
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
10,
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers (78). (1328).
FT CDS Location/Qualifiers
source 1..2716
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 46.8%; Score 1260.4; DB 2; Length 2716;
Best Local Similarity 79.7%; Pred. No. 0;
Matches 1716; Conservative 0; Mismatches 376; Indels 60; Gaps 17;
QY 594 TCCTGAAGAATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAGCGGGCTA 653
DB 1 TACTAAAGAATGAGAAAGATGAGGTGCAAAATTTACAAAATATCATTCGAAGTCAAGCTA 60
QY 654 CTCAGTATAATCATGATGTGAAGAGAGAGAGCGTGAATATAATAAGCTAAAGAGAGCGCC 713
DB 61 CTCAGTATAATCATGATATGAAGAGAAAGAGCGTGAATATAATAAAGCTGAAGGAAAGTGC 120
QY 714 TGCATCAGCTCGTTATGAACGAAGAGGATATAAATCATAGCCATGATGTTTAAATATG 773
DB 121 TACATCAACTTGTATGAACGAAGAGATGAAGAAATAGCTATGAGCAATTTTGAATATG 180
QY 774 TGGGTGAGCTGTGCGCAACGAGCTCATCGAGGACTGACAAAACAGAACGAGGAGATG 833
DB 181 TCGGAGAGCTGATGGAAAGAGAGCTCTCGAGGAGCTGGTAAAGCTGAAGCCAGGATG 240
QY 834 AAGATGAGATGTACAAAATTCCTTTGAATGATTATGATACCGCCAGAGAGAGATCCTGA 893
DB 241 AAGATGAATGTATAAATTCCTTTGAATGATTATGATATCGTCAGAAAACAAATCCTAA 300
QY 894 TGGAGAACCGGAGCTGAAGAAGTCTCTCAGCAGATGAAGAAGAGATGATCTCTCC 953
DB 301 TGGAAATGCGAACTTAAGAAGGTCTTCAACAAATGAAAAGGAAATGATTTCTCTTC 360
QY 954 TGTCTCCTCAGAGAAGAGCCAGGAAAGAGAGAGAGCGGACAGGACTGTTTGCTA 1013
DB 361 TTTCTCCCCAAAGAAAGAAACCTAGAGAAAGATGATAGTACAGGAAGTGT--TA 417
QY 1014 TCTCCGATATAGAAGATGACTCTGGGAACTGAGCAGAGACAGCGTGTGGGGCCTTTCCCT 1073
DB 418 TTTCCGATGTTGAAGAAGATGCGGGGAACCTAAGCAGAGAGATGATGTGGGACCTTTCTCT 477

QY 1074 GTGACACTGTGAGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTGAAAA 1133
DB 478 GTGAAACTGTGAGAGAGCAGCTTACAAACAGCATCAGAAAACAGTGGAGAAATTTGAAAA 537
QY 1134 GTCATGTAGAAAAAATCGATAAACCAAGCTTCGAAAGTACACTCAGAGGGCCTTAATGAGG 1193
DB 538 GTCATGTAGAAAAGCTGTGATAACCAAGTTTCAAAGGTACACCTGGAAAGGTTTTAATGATG 597
QY 1194 AGGACGTCTCTCAAGCAAGACCATGAGCAAGAGACTGAGAAAACCTGGAGCTGAGATTG 1253
DB 598 AAGATGTAAATCTCAGCAAGAGCATGAACTGAAAACCTGAGTTAGAAATTC 657
QY 1254 AGCGGTGTAAGAGATGATCAAGGCTCAGCAGAGCTCTTACAGCAGCAGCTGCCACCA 1313
DB 658 AGCAGTGTAAAGAAATGATTAAGAACTCAGCAAGAGCTTTTACAGCAGCAGCTCCTACTG 717
QY 1314 CGTGTGATGATGACACCACCTCAGCTGTGGGAGACTGTTACTTGTGGAAGAAAGAAC 1373
DB 718 CATATGATGATGATACCACTTCACATTACGAGACTGTTATTGTTGGAAGAAAGAAC 777
QY 1374 GCCTTAAAGAGAGTGGACCCCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGGC 1433
DB 778 GTCTCAAAGAGAAATGCTCCCTTTTAAAGAGCAGAAAAGAAATTTTGAGAGGAGAGAC 837
QY 1434 GAAGCTTTACAGAGCTGCCATTCGATTGGGTTGGAGAGAAAGGCGTTTTTGAAGAAAGAC 1493
DB 838 GAAGCTTTACAGAGCGCGCTATTTCGCTGGGATTTGGAGAGAAATGGCAATTTGGAAGAGAA 897
QY 1494 GAGCAGCTGGGTAAGAGCAGCAGTTTTTAAACATGACGAACTTTGACCAACAGAACTCAG 1553
DB 898 GAGCCAGTTGTTAAAGCAGCAGTTTTCTAAATATGACTACCTTTGACACCAGAACTCAG 957
QY 1554 AAAATGTGAAACTTTTCAGTGCCTTCTCAGAGAGTTCTGATCCAGACAACTTTATAGTCC 1613
DB 958 AAAATGTGAAACTTTTCAGTGCCTTCTCAGAGAAATTTCTGATTTGGGACAACTTTATAGTGC 1017
QY 1614 ACTCAGCGCCACGCAAAAAGAGCTACACAGTGGGCTAAATGGGTTGCCAGTTTGCACAT 1673
DB 1018 ACTCAGGAGCGCCGCAAAAAGAGCTCAGAGTGTCTATGGTCTCCAGTTTGCATGT 1077
QY 1674 CAAAACCTGACTAAATCTCTTCCTGCCCTCACCTTCTACTTCTCAGACTTTTGGCAGACACATT 1733
DB 1078 CTAACCTTACTAAATCTCTTCTCCTGCTTCCACTTCCAGACTTTTGCAGACAGCTT 1137
QY 1734 CATGTGTCTGAACACAGTTCCATCAGTGTCTGTAATATACTCCTCGAAGAAAGTAAAC 1793
DB 1138 CCTGCATATCTGAACATAGTTCAATCAATGTACTGAATATAAAGTGTGAAAGAAATTAAC 1197
QY 1794 CAAAGTGAAGTTTGAAGAGAAAGCAGCGATCAGAAAGTGGAGCTGAGTCCAGAGGCCAGCT 1853
DB 1198 CAAATCAGTTTGGAGAGAAATGTACAAATCAAAAATGGAGTGTGGCGTCAAGACTGGAT 1257
QY 1854 CGCGGGAGGGGTGCTACAGCGGATGCTCCTCGGCCCTTTCAGAGCGCTCACGGGAGCCGAG 1913
DB 1258 CACAGGAAGGTTGCTATAGTGGATGCTCCTTGAGCTACACAAATCTCATGTAGAAAAAG 1317
QY 1914 ATGACTTACCTTAAATGTGGGGCTGAGTGTCTTCCAGATGTGGCTAGAGG----- 1968
DB 1318 ATGACTTACCTTAGACATGTGGACTGGAAATTTTTTCAATTAATGTGTTTCATCAAGTTTCA 1377
QY 1969 -----AGTTTGACACAGGGGTGTAGCATAAAAGTCAAGTGGTCT-----TAACTTTAAGATGCTC 2017
DB 1378 CATCTAAGTTTGAACACAGGGTGTGTCATAAAGTCAAGTTATCTCTAATAACTTAAGATGCTC 1437
QY 2018 AGAGTTGCTTTGTTGGACTTTCCTGCTCTTCCCCCAAGAGAGCTGAAATGCT--AAGCTACTT 2076
DB 1438 TGAGTTGCTTTGTTGGACTTTCCTGCTCTTCCCCCAAGAGAGTTGAAATCTTAAATCTATTT 1497
QY 2077 AAAAGGATGCAAGACTTTGG-----TTGTGTGTTAGTAACAGAGCCCTGCTCTGTGAC 2132
DB 1498 AAAAGGATATAAAGACTTTGGATATGTAATTTTTTAGTAACAGAGCATCTGGTTCTGTGA 1557

```
QY 2133 TGCAGGAATGCAT-GGCCTTTGGATGGAACAGAGCGCTGGAATGA--TTGCCTCGCCA 2189
Db 1558 TAAAGGAATGTAGATGTTTGATGGAACAAAGACACTAGACGAGTTCTCTTTATA 1617
QY 2190 GGTACCGAGAGAGACACTTTTAGGAGCTGTTCTGTAAACATTAATATTCGTCCTCAAG 2249
Db 1618 GGTATTAATAATAGCACCTTTTAGGAAACTGATTATTTGTAATGTTAAATTTTGCTCATA 1677
QY 2250 TCGTGTGTCATGGAAGTGTAGCCTTTTACCTTGAATGTATATCTAGATTTTAAACAAG 2309
Db 1678 TATAGTTGTCATGGAAGTTTAGCCTTTTACTTTGAATGTATATCTAGATTTTAAACAAG 1737
QY 2310 CAGGTTCTATATTTATTTATGTTTGTAGTGTATTTGGATTACCTCTTTCATATGTTTT-- 2367
Db 1738 CGAGTTCTATATTTATTTATGTTTGTAGTGTG-GTTTGAATTAACCTCTTTCATATGTTTAA 1796
QY 2368 -----GTCTCTGTACATAAATATACATGACATATGTTAAAGAGCTTT 2408
Db 1797 ATAAAGTGAATTTATGTATGTTTGTACATAGATACATGATTATGTTAAAGAGCTTT 1856
QY 2409 AAGTTTAAACACTTCACACCATGCTTGATGTATAGCATTTTCATGCCAATTAATATGTTTT 2468
Db 1857 AAGATTTAAAGTTTTCACA-CAACCATAGTATAGTATTTTCATGCC-AGTAAATTTTTT 1914
QY 2469 CAGTGGCATGGTGTACAGAGG--TTAGGACCATGCCACATGACAGTTAAGACTTTAT 2526
Db 1915 TAGTGGTATCTGTTTACAGATGTATAGGACCATGTATGATTAACATTTAAGA-ATTCT 1973
QY 2527 TTTTAAAGCCATCGGCAATAAATAATCAAGCCCTTCAT-AAGCTGAGTTC---AGAT 2582
Db 1974 CTTTAAATACATCGGCAATAAATTTGAAAGTATTCATGAACTGAGTCTTTTAGAT 2033
QY 2583 AACTAGAACTACTAAGTTTACATTTTGTAGATTTTAAAGCAAT-GTATTTTATTTTATA 2641
Db 2034 AATCAACACTACTAACAATTTTACATTTTGTAGATTTTATGACATTTAGATTTTATTTGTA 2093
QY 2642 TATGT-GAATGTATTAATTTCTAAGAGGAATATTCATTATGAGTAATGGG 2692
Db 2094 TATGTAGAAATTTAATTTTAAAGGACTATTGATGATAGAAATAGG 2145

RESULT 14
AX877359
LOCUS AX877359 2716 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 12264 from Patent EP1074617.
ACCESSION AX877359
VERSION AX877359.1 GI:40032095
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1
Ota, T., Isoigai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
Primers for synthesising full-length cDNA and their use
Patent: EP 1074617-A 12264 07-FEB-2001;
Research Association for Biotechnology (JP)
Location/Qualifiers
1. 2716
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
78..1331
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE89690.1"
/db_xref="GI:40032096"
/translation="MKREREYNKLERLHOLVNMKKDKKIAMDILNVGRADKRGK
WRGTGEARNDENVKILNDYEVROKQILMENAELKKVLQOMKEMISLLSPKKKP
RERVDDSTGTVTVEEDAGELSRSMWDLSCETVRSLTNSIKQWRLKSHVEKLD
NQVSKVHLEGFNDVEDVISRQDHEBETEKLELEIQOCKEMIKTQOOLQOOLATAYDD
```

ORIGIN

```
Query Match 46.8%; Score 1260.4; DB 2; Length 2716;
Best Local Similarity 79.7%; Pred. No. 0;
Matches 1716; Conservative 0; Mismatches 376; Indels 60; Gaps 17;

QY 594 TCCTGAAGAATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCAGCCGGCTA 653
Db 1 TACTAAAGATGAGAAAGATGAGGTGCAAAAATTTACAAAATATCATTTGCAATCGAGCTA 60
QY 654 CTCAGTATAATCATGATGTGAAGAGAGAGCGTGAATATAATTAAGCTAAAGAGAGCGCC 713
Db 61 CTCAGTATAATCATGATATGAAGAAAGAGCGTGAATATAATTAATTAACCTGAAGAAACGTC 120
QY 714 TGCATCAGCTCGTTATGAACAAGAAAGGATAAAACATAGCCATGGATGTTTTTAAATATG 773
Db 121 TACATCAAATTGTTATGAACAAGAAAGATAGAAAATAGCTATGGACATTTTTTGAATATG 180
QY 774 TGGGTTCGAGCTGATGGCAACAGAGCTCATTCGAGGACTGACAAAACAGAAAGCCAGAAATG 833
Db 181 TCGGGAGAGCTGATGGAAGAAAGAGGCTCTCGAGGACTGGTAAACTGAAGCCAGAAATG 240
QY 834 AAGATGAGATGTACAAAATTTCTGTTGAATGATTAGTACCGCCAGAGAGAGATCCTGA 893
Db 241 AAGATGAATGTATAAATTTCTTTGAATGATTATGATATCGTCAGAAAACAAATCCTAA 300
QY 894 TGGAGAACCGGAGCTGAAGAAAGGTCCTCAGAGAGATGAAGAAGAGATGATCTCTCTCC 953
Db 301 TGGAAAATGCAAGAACTTAAGAAAGGTTCTTCAACAAATGAAAAGGAAATGATTTCTCTTC 360
QY 954 TGTCTCTCAGAGAAGAAAGCCAGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1013
Db 361 TTTCTCCCAAGAGAAAGAACTAGAGAAAGAGTAGATGATAGTACAGAACTGT---TA 417
QY 1014 TCTCCGATATAGAAGATGACTCTGGGGAATCTGAGCAGAGAGAGAGAGAGAGAGAGAGAG 1073
Db 418 TTTCCGATGTTGAAGAGATGCCGGGAATCAAGCAGAGAGAGATGATGCGGACCTTTCTCT 477
QY 1074 GTGACATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1133
Db 478 GTGAAACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 537
QY 1134 GTCATGTAGAAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGAGAGAGAGAGAGAG 1193
Db 538 GTCATGTAGAAAAAGCTTTGATAACCAAGTTTCAAAGGTACACCTGGAAGGTTTTTAATGATG 597
QY 1194 AGGAGCTCATCTCAGCAAGACCATGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1253
Db 598 AAGATGTAATCTCAGCAAGACCATGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657
QY 1254 AGCGGTGTAAAGAGATGATCAAGGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1313
Db 658 AGCAGTGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717
QY 1314 CGTCTGATGATGACACACCTCAGTGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1373
Db 718 CATATGATGATGATACCACTTCACTATTAACCAAGTTTCAAAGGTACACCTGGAAGGTTTT 777
QY 1374 GCCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1433
Db 778 GTCTCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 837
QY 1434 GAAGCTTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1493
Db 838 GAAGCTTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897
QY 1494 GAGCAGCTGGTAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1553
Db 898 GAGCAGCTGGTAAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957
```

Qy 1554 AAAATGTGAAACTTTTTCAGTCCCTTCTCAGGAAGTTCTGATCCAGACAATCTTATAGTCC 1613
Db 958 AAAATGTGAAACTTTTTCAGTCCCTTCTCAGGAAGTTCTGATGGGACAATCTTATAGTGC 1017
Qy 1614 ACTCAGGCCACGGCAAAAGAGCTACACAGTGTGGCTTAATGGGGTGCACGCTGCGACAT 1673
Db 1018 ACTCGAGCGCAGCAAAAGAGCCTCACAGTGTGCTAATGGGTCTCCAGTTTGCATGT 1077
Qy 1674 CAAAACTGACTAAATCTCTTCTGCTGCTCACCCTTCTACTTCAGACTTTTCGCCAGACACAT 1733
Db 1078 CTAACCTTACTAAATCTCTTCTGCTGCTCACCCTTCCACTTTCAGACTTTTTCGCCAGACAGCT 1137
Qy 1734 CATGTGCTCTGAAACACAGCTCCATCAGTGTGCTGAAATATACCTCCTGCAAGAAAGTAAAC 1793
Db 1138 CCGTCATATCTGAAACATAGTTCAATCAATGACTGAAATATACCTGCTGAAGAAATTAAC 1197
Qy 1794 CAAGTGAGGTTGCAAGAGAAACGCGATCAGAAAGTGGAGCGTGCAGTCGAGGCCAGCT 1853
Db 1198 CAAATCAGGTTGGAGGAGAATGTACAAATCAAAATGAGTGTGGCGTCAAGACCTGGAT 1257
Qy 1854 CGCGGAGGGGTGCTACAGCGGATGCTCCTCGGCCTTCAGGAGCGCTCAGGGGACCGAG 1913
Db 1258 CACAGGAAGGTTGCTATAGTGATGCTCCTTGGAGCTACACAAATCTCATGTAGAAAAG 1317
Qy 1914 ATGACTTACCTTAATGTGCGGGCTGCAGTCTGTTCCAGATGCGCTAGAGG---- 1968
Db 1318 ATGACTTACCTTAGACATGTGGAGTGGAAATTTTTTTCATTAATGTGTTCACTAGTTTCA 1377
Qy 1969 -----AGTTGACACAGGGTGTAGCATAAAGTCAGTCCGTC-----TAACCTTAAGATGCTC 2017
Db 1378 CATCTAAGTTGAAACAGGGTGTGTCATAAAGTCAGTTATCTCTAATAACTTAAGATGGTC 1437
Qy 2018 AGAGTTGTTTGTGGACTGCTGCTCTTCCGCCAAAGAGCTGAAATGCT--AAGTACTT 2076
Db 1438 TGAGTTGTTTGTGGACTGCTCCTGCTCTTCCGCCAAAGAGTGAATCTTAATCTATTT 1497
Qy 2077 AAAAGGATGCAAGCTTTGG-----TTGTGTGTTAGTACAGAACCCCTGCTCTGTGAC 2132
Db 1498 AAAAGGATATAAAGCTTTGGATATGTATTTTTTTAGTAAACAGAGCATCTGGTCTGTGAA 1557
Qy 2133 TGCAGGAATGCAT--GGCCTTTGGATGGAACAGAGCCCTGGAATGA--TTGCCCTCGCCA 2189
Db 1558 TAAAGGAATGTATAGATGTTTGGATGGAACAAAGACACTAGACTGAGTTTCTCTTTATA 1617
Qy 2190 GGTACCGAGAGAGCAGCTTTTAGGACTGGTTCCTGTGTAACATTAATATTCGTTCCCAAG 2249
Db 1618 GGTATTAATAAATAGCAGCTTTTAGGAACACTGATTATTTGTAATGTTTAAATTTTGTCTCAA 1677
Qy 2250 TGTGTTGGCATTGGAAGTGTAGCCTTTTACTTGAATGTATCTGTAGATTTTAAACAAAG 2309
Db 1678 TATAGTTGGCATTGGAAGTTTAGCCTTTTACTTGAATGTATCTGTAGATTTTAAACAAAG 1737
Qy 2310 CAGGTTCTATATTTATTTAGTTAGTGTGATTTTGGGATTAACCTCTTTTCATATGTTTT-- 2367
Db 1738 CGAGTTCTATATTTATTTAGTTAGTGTG--GTTTGAATAATTAACCTCTTTTCATATGTTTTAA 1796
Qy 2368 -----GTGCTGTGTACATAAATATATACATGACTATGTTTAAGAGGCTTT 2408
Db 1797 ATAAGTGAAATTTATGTATGTTTGTGTACATAGATACATGATTATGTTTAAGAGGCTTT 1856
Qy 2409 AAGGTTTAAAAACTTCACACCATGCTTCAGTATAGCATTTTCATGCAATTAATAATGTTTT 2468
Db 1857 AAGATTTAAAGTTTCACA--CAACCATAAAGTATAGTATTTTCATGCC--AGTAAATTTTTT 1914
Qy 2469 CAGTGGCATGTTGTTTACAGAGG--TTAGGACCACTGCCACATGACAGTAAAGACTTTAT 2526
Db 1915 TAGTGGTATCTGTTTACAGATGTATTAGGACCACTTGTATGATTTAAGA--ATTCT 1973
Qy 2527 TTTTAAAGCATCTGGGCAATAAATTCAAAGCCCTTCAT--AGCTGAGTTC---AGAT 2582
Db 1974 CTTTAAATCATCTGGGCAATAAATTTGAAGGTTATTCCTCAAGAGCTGAGTTCTTTAGAT 2033

Qy 2583 AACTAGAACTACTAAACGTTACTATTTTGTAGATTTTTTAAAGCAAT--GTATTTTATTTTATA 2641
Db 2034 AATCAACACTACTAAACATTTACATTTTGTAGATTTTTTATGACATTTAGATTTTATTTCTA 2093
Qy 2642 TATGT--GAATGTTATTAATTTCTTAAGAGGAATATTTGATTATGGAGTAATGGG 2692
Db 2094 TATGTAGAATATTATTAATTTTAAAGGACTTATGATGATAGAGAATAGGG 2145
RESULT 15
AK001710 2716 bp mRNA linear PRI 20-JAN-2006
LOCUS Homo sapiens cDNA FLJ10848 fis, clone NT2RP4001407, highly similar
to Homo sapiens mRNA for KIAA0923 protein.
AK001710
AK001710.1 GI:7023139
VERSION oligo capping; fis (full insert sequence).
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1
Oca,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,
Wakamatsu,A., Hayaashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,
Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,
Nagahara,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagaesuma,M.,
Shiratori,A., Sudo,H., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,
Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K.,
Tanihara,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,
Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,
Hirooka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,
Yosida,M., Hotuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,
Hara,H., Tanase,T.O., Nomura,Y., Togiya,S., Komai,F., Hara,R.,
Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,
Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichiara,T.,
Shibata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,
Terahima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,
Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,
Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,
Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,
Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,
Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,
Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T.,
Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K.,
Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T.,
Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,
Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,
Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R.,
Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
Complete sequencing and characterization of 21,243 full-length
human cDNAs
Nat. Genet. 36 (1), 40-45 (2004)
1470239
2
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,
Tanai,H., Kimata,M., Watanabe,M., Hirooka,S., Ishii,S., Kawai,Y.,
Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K.,
Masuho,Y. and Kanehori,K.
NEDO human cDNA sequencing project
Unpublished
3 (bases 1 to 2716)
Isogai,T. and Otsuki,T.
Direct Submission
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: flj-cdna@nifty.com, Tel.81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library

construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES

Location/Qualifiers
1..2716
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="NT2RP4001407"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP4"
/note="cloning vector: pME18FL3
mRNA from NT2 neuronal precursor cells after 2-weeks
retinoic acid (RA) induction."

ORIGIN

Query Match 46.8%; Score 1260.4; DB 5; Length 2716;
Best Local Similarity 79.7%; Pred. No. 0; Mismatches 376; Indels 60; Gaps 17;
Matches 1716; Conservative 0;

594 TCCTGAAGAATGAGAAAGATGAGGTACAAAAATTACAAAATATCATAGCCAGCCGGCTA 653
Db 1 TACTAAAGATGAGAAAGATGAGGTGCAAAATTACAAAATATCATTCGAGTCCAGCTA 60

654 CTCAGTAAATCATGATGTAAGAGGAAGGAGCGTGAATATTAATTAAGCTAAAGAGCGCC 713
Db 61 CTCAGTAAATCATGATGTAAGAGGAAGGAGCGTGAATATTAATTAAGCTAAAGAGCGTC 120

714 TGCATCAGCTCGTTATGAAAGAGGATTAACATAGCCATGATGTTTAAATATG 773
Db 121 TACATCAACTTGTATGAAACAGAAAGATAGAAAATAGCTATGACATTTTGAATATG 180

774 TGGTTCGAGCTGATGGCAACGAGGCTCATGGAGGACTGACAAAACAGAACGACGGAATG 833
Db 181 TCGGAGAGCTGATGAAAGAGAGGCTCTCGAGGACTGGTAAACTGAGCCAGGAATG 240

834 AAGATGAGATGACAAAATCTCTTGAATGATATGATGACCGCAGAGAGAGATCTCTGA 893
Db 241 AAGATGAAATGTATAAAATCTCTTGAATGATATGATGATATCGTCAGAAAACAAATCTCTTA 300

894 TGGAGACCGGAGCTGAAGAGGCTCTCCAGCAGATGAAGAGAGATGATCTCTCC 953
Db 301 TGGAAATGCGAGAACTTAAGAAAGGTTCTCAACAAATGAAAAGGAAATGATCTCTTC 360

954 TGTCTCTCTCAGAAAGAACCCAGGAAAGAGCAGAGAGCGCACGCACTGTGTGCTA 1013
Db 361 TTTCTCCCAAGAGAGAAACCTAGAGAAAGTAGATAGTACAGGAATGTT--TA 417

1014 TCTCGATATAGAGATGACTCTGGGAACTGAGCAGAGACAGCGTGTGGGCTTTTCT 1073
Db 418 TTTCCGATGTTGAAGAAGATCCGGGAACTAAGCAGAGAGATGATGTGGGACCTTCT 477

1074 GTGACATGTGAGAGAGAGCTGACAAACAGCATCAGAAACAGTGGAGATTTTGA 1133
Db 478 GTGAACCTGTGAGAGAGAGCTTCAACACAGCATCAGAAACAGTGGAGATTTTGA 537

1134 GTCATGTAGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGGCTTAAATGAGG 1193
Db 538 GTCATGTAGAAAGCTTGTAAACCAAGTTTCAAGGTACACTCGAAGGTTTAAATGATG 597

1194 AGGAGCTCATCTCAGCAAGACCATGAGCAAGAGACTGAGAACTGGAGCTGGAGATTG 1253
Db 598 AAGATGTAATCTCAGCAAGACCATGAAACAGAACTGAAAACTCGAGTTAGAAATTC 657

1254 AGCGGTGAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGCCACCA 1313
Db 658 AGCAGTGAAGAAATGATTAATACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTG 717

1314 CGTGTGATGATGACACCACTCTCTGTTGGCAGACTGTTTCTCTCGTGAAGAAAGAAC 1373
Db 718 CATATGATGATGATACCACTTCACTATTACAGACTGTTTATTTGTTGGAAGAAAGAAC 777

1374 GCCTTAAAGAAAGAGTGGAGCCCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGC 1433
Db 778 GTCTCAAAGAAAGATGGTCCCTTTTAAAGAGCAGAAAAAGAAATTTTGAGAGGGAGAC 837

1434 GAAGCTTTACAGAAAGCTGCCATTCCGATTTGGGGTTGGAGAGAAAGCGTTTGAAGAGAGC 1493
Db 838 GAAGCTTTACAGAAAGCGCTATTCCCTCGGATTTGGAGAGAAATGCAATTTGAAGAGAAA 897

1494 GAGCAGCTGGGTAAAGCAGCAGTTTTTAAACATGACCACTTTTCCACCAAGAACTCAG 1553
Db 898 GAGCAGTTGGTTAAAGCAGCAGTTTTCTAAATATGACTACCTTTGACCAAGAACTCAG 957

1554 AAAATGTGAATCTTTTCAAGTCCCTTCTCAGGAAGTTCTGATCCAGCAATCTTATAGTCC 1613
Db 958 AAAATGTGAATCTTTTCAAGTCCCTTCTCAGGAAGTTCTGATTTGGGACAATCTTATAGTGC 1017

1614 ACTCAGCCGACGCGCAAGAAAGCTACACAGTGGGCTTAATGGGTGCCAGCTTGCACAT 1673
Db 1018 ACTCAGGCGCCGCGCAAGAAAGCTCACAGTGTCTTAATGGGTCTCCAGTTTGCATGT 1077

1674 CAAAACCTGACTAAATCTCTCTGCTCACCTTCTTACTTTCAGACTTTTCGCGAGACACATT 1733
Db 1078 CTAAACTTACTTAATCTCTCTGCTCACCTTCCACTTCAGACTTTTTCGCGAGACAGCT 1137

1734 CATGTGTCTGAAACACAGTTCCATCAGTGTCTGAATATACTCTCTGAAGAAAGTAAAC 1793
Db 1138 CCTGCATATCTGAACATAGTTCAATCAATGACTGAATATAAATGCTGAAGAAATTAAC 1197

1794 CAAGTGAAGTTGCAAGAGAAAGCAGGATCAGAAAGTGAGGCTGAGTGGAGCCAGCT 1853
Db 1198 CAATCAAGTTGGAGGAGAAATGTAACAATCAAAAATGGAGTGTGGCGTCAAGACCTGGAT 1257

1854 CGCGGAGGGGTGCTACAGCGGATGCTCTCGGCTTCAGGAGCGCTCACGGGACCCAG 1913
Db 1258 CACAGGAAGTTGCTATAGTGAGTCTCTTGGAGCTCAAAATTTCTCATGTAGAAAAG 1317

1914 ATGACTTACCTTAAATGTGGGGTGCAGTGTCTGTCCAGATGTGCGCTAGAGG---- 1968
Db 1318 ATGACTTACCTTAGACATGTGGAATGGAATTTTTCATTAATGTGTTTCATCAAGTTTCA 1377

1969 -----AGTTGACACAGGCTGTAGCATAAAGTCAAGTCTGTC-----TAACTTAAAGTCTC 2017
Db 1378 CATCTAAGTTGAAACAGGGTGTGTATAAAGTCAAGTTATCTCTAATAAATTAAGATGTC 1437

2018 AGAGTGTGTTTGGGACTTGGCTGTCTTCCCAAGAGCTGAAATGCT--AAGTACTT 2076
Db 1438 TGAGTGTGTTTGGGACTTCCCTGTCTTCCCAAGAGAGTTGAATCTTAAATCTATTT 1497

2077 AAAAGGATGCAAGCTTTGG-----TTGTGTGTTTAAAGAGCCCTTGGCTCTGTGAC 2132
Db 1498 AAAAGGATATAAAGCTTTGGATATGATTTTAAAGTAAAGCAATCTGTTCTGTGAA 1557

2133 TGCAGGAATGCAT--GGCGTTTGGATGAAACAGAGCGCTGGAATGA--TTGCTCTGCCA 2189
Db 1558 TAAAGGAATGTATAGATGTTTGGATGGAACAAAGCACTAGACTGAGTTTCTCTTATA 1617

2190 GGTACCCAGAGAGACCTTTTAGGCACTGGTTCCTGTAAACATTAATAATTTGCTCCCAAG 2249
Db 1618 GGTATTAATAATAGCCTTTTAGGAACTGATTAATGTAATGTTTAAATTTTGTCTCAA 1677

2250 TGTGGTTGGCATTTGGAAGTGTAGCCTTTTACTTTGAATGTATATCTGTAGATTTTAAACAAAG 2309
Db 1678 TATAGTTGGCATTTGGAAGTTTGGCTTTTACTTGAATGTATCTGTAGATTTTAAACAAAG 1737

2310 CAGTTCTATATTTATTTATGTTAGTGTGATTTTGGGATTTACCTCTTTCATATGTTTT-- 2367
Db 1738 CGAGTTCTATATTTATTTATGTTTAGTGTG--GTTTGAATTTACCTCTTTCATATGTTTTAA 1796

2368 -----GTGCTCTGACATAATATACATGACTATGTTTAAAGAGGCTTT 2408
Db 1797 ATAAAGTGAATTTATGATGTTTGTATACATGATGATTAATGTTTAAAGAGGCTTT 1856

Qy	2409	AAGGTTTAAAACTTCACACCATGCTTGAGTATAGCATTTTCATGCCAATTAAAAATGTTTT	2468
Db	1857	AAGATTTTAAAGTTTCACA-CAACCATAAAGTATAGTATTTTCATGCC-AGTAAAAATTTTT	1914
Qy	2469	CAGTGGCATGGTGTGTTTACAGAGG--TTAGGACCACTGCCACATGACAGTTAAGACTTTAT	2526
Db	1915	TAGTGGTATTCTGTTTACAGATGTATTAGGACCATTGATGCATTACATTTAAGA-ATTCT	1973
Qy	2527	TTTTAAGCCATCTGGGCAATAAATTCAAAGCCCTTCAT-AAGCTGAGTTC--AGAT	2582
Db	1974	CTTTAATACATCTGGGCAATAAATATTGAAAGGTATTCATGAAGCTGAGTCTTTAGAT	2033
Qy	2583	AACTAGAACTACTAACGTTACATTTTGAGATTTTAAAGCATT-GTATTTTATTTTATA	2641
Db	2034	ATCAACACTACTAACATTACATTTTGAGATTTTATGACATTAGATTTTATTTTGTGA	2093
Qy	2642	TATGT-GAATGTTATAATTTCTAAGAGGAATATTGATTCGAGTAATGGG	2692
Db	2094	TATGTAGAATATTATAATTTTTTAAAGGACTATTGATGATAGAAGAATAGGG	2145

Search completed: June 14, 2006, 00:53:15
Job time : 14790 secs

GenCore version 5.1.1.9
 Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2006, 20:46:55 ; Search time 11414 Seconds
 (without alignments)
 13188.614 Million cell updates/sec

Title: US-10-644-084-1
 Perfect score: 2692
 Sequence: 1 cgtaggagagtcacaggagc.....ttgattatggaggaataatgggg 2692

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : EST:
 1: gb_est1:
 2: gb_est3:
 3: gb_est4:
 4: gb_est5:
 5: gb_est6:
 6: gb_hc:
 7: gb_est2:
 8: gb_est7:
 9: gb_est8:
 10: gb_est9:
 11: gb_ges1:
 12: gb_ges2:
 13: gb_ges3:
 14: gb_ges4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2676	99.4	3185	6 AK043865	AK043865 Mus muscu
2	2676	99.4	3413	6 AK031356	AK031356 Mus muscu
3	1985.4	73.8	2245	6 AK077112	AK077112 Mus muscu
4	1973	73.3	2371	6 AK049080	AK049080 Mus muscu
5	1832	68.1	1845	14 AY412493	AY412493 Mus muscu
6	1701.4	63.2	3154	6 BC035580	BC035580 Homo sapi
7	1701.4	63.2	5658	6 HS0801317	AL133046 Homo sapi
8	1416.6	52.6	2111	6 CR613292	CR613292 full-leng
9	1363.8	50.7	1845	14 AY412491	AY412491 Homo sapi
10	1357.4	50.4	1845	14 AY412492	AY412492 Pan trogl
11	876.4	32.6	878	9 CX565842	CX565842 UI-M-IB0-
12	812	30.2	840	5 CF949831	CF949831 UI-M-H20-
13	810	30.1	834	9 CX562093	CX562093 UI-M-FC0-
14	802	29.8	804	8 CN539511	CN539511 UI-M-H00-
15	798	29.6	798	9 CX562859	CX562859 UI-M-IB0-
16	789	29.3	822	8 CV563162	CV563162 UI-M-FC0-
17	768.4	28.5	771	3 B0179212	B0179212 UI-M-EW0-
18	761	28.3	933	4 CB205832	CB205832 AGENCOURT
19	753.2	28.0	771	5 CF737129	CF737129 UI-M-HD0-

20	750.2	27.9	770	5	CF729394	CF729394 UI-M-HD0-
21	749	27.8	760	8	CN528686	CN528686 UI-M-HQ0-
22	748.2	27.8	845	3	BQ769157	BQ769157 UI-M-FC0-
23	739.2	27.5	746	8	CN528831	CN528831 UI-M-HQ0-
24	721.4	26.8	723	5	CF737957	CF737957 UI-M-HQ0-
25	710.4	26.4	712	5	CF732776	CF732776 UI-M-GZ0-
26	690	25.6	736	8	CX226608	CX226608 MBM02133
27	679	25.2	679	8	CN532570	CN532570 UI-M-HQ0-
28	677.8	25.2	683	8	CO424815	CO424815 UI-M-HQ0-
29	675	25.1	687	9	CX568194	CX568194 UI-M-IB0-
30	658	24.4	903	10	DV893190	DV893190 LB02726.C
31	653.2	24.3	1005	1	AL582610	AL582610 AL582610
32	643	23.9	643	2	BG085173	BG085173 H3109D09-
33	638	23.7	640	5	CF737822	CF737822 UI-M-HD0-
34	633	23.5	833	8	CV117234	CV117234 AGENCOURT
35	623.4	23.2	638	8	CN720771	CN720771 E0809F12-
36	622.6	23.1	954	1	AL561392	AL561392 AL561392
37	622.2	23.1	635	5	CF899025	CF899025 A024TH06-
38	616.4	22.9	860	10	DT831879	DT831879 LB00413.C
39	614.4	22.8	633	1	AV279314	AV279314 AV279314
40	614.2	22.8	805	2	BG242849	BG242849 602355829
41	614.2	22.8	828	10	DV803752	DV803752 LB01326.C
42	612.2	22.7	658	7	B8614736	B8614736 BB614736
43	606.8	22.5	623	4	CA548100	CA548100 C0801B11-
44	601.2	22.3	752	7	BF539334	BF539334 602049514
45	600	22.3	600	2	BG804175	BG804175 0252-45 M

ALIGNMENTS

RESULT 1
 AK043865
 LOCUS
 DEFINITION Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched library, clone:A830043F14 product:HYPOTHETICAL 71.0 KDA PROTEIN homolog [Mus musculus], full insert sequence.

ACCESSION AK043865
 VERSION AK043865.1 Gi:26335971
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE
 1 Carninci P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 2 10349636
 3 Carninci P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 11042159

AUTHORS
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 Sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 11076861

TITLE
 JOURNAL
 PUBMED
 REFERENCE
 4 The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection

JOURNAL REFERENCE AUTHORS	Nature 409, 685-690 (2001)
TITLE	The FANTOM Consortium, the RIKEN Genome Exploration Research Group Phase I and II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL REFERENCE AUTHORS	Nature 420, 563-573 (2002)
TITLE	RIKEN Genome Exploration Research Group, Genome Science Group (Genome Network Core Team) and the FANTOM Consortium. Antisense Transcription in the Mammalian Transcriptome
JOURNAL REFERENCE AUTHORS	Science 309, 1559-1563 (2005)
TITLE	The FANTOM Consortium, Riken Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). The Transcriptional Landscape of the Mammalian Genome
JOURNAL REFERENCE AUTHORS	Science 309, 1559-1563 (2005)
TITLE	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hirozane, T., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
JOURNAL REFERENCE AUTHORS	Submitted (16-JUL-2001) Yoshihide Havaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/ , Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers 1. 3185 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:A830043F14" /db_xref="taxon:10090" /clone="A830043F14" /tissue_type="cortex" /clone_libs="RIKEN full-length enriched mouse cDNA library" /dev_stage="10 days neonate" 389. .2233
CDs	/note="unnamed protein product; HYPOTHETICAL 71.0 KDA PROTEIN homolog [Mus musculus] (SPT AAH21749, evidence: FASTY, 99.8%ID, 100%length, match=1842) putative" /codon_start=1 /protein_id="BAC31684.1" /db_xref="GI:26335972" /translation="MGDMVTVDVLTENKLSQYTSKMSPPSLYSQVLCSVP LSKNVQVGFVCTGEISQISYLDQELTTFGPFSLYESKSEAKRELNIIVALNC MNEILLVKNLQAESVETONLQSDMDHLQSCYAKLKEOLTSRRMIGLQERD QLOCKNSLHOLLKXEDVQKLNIIASRATOVNHDVSKEREVYKLERHLQVNV KDKNIAMDVLNYVGRADGRKGSWRTEARNEDNKKYLLNDYERQQLMENE LKKVLQMKMEISLSPQKKPRERADGTGTVAISDIEDDSGELSRSVWGUSCDT VREGTNSIRKQWILKSHVEKLDNQASKVHSEGLNEEDVISRQDHOETEKELEIE RCKEMIKAQQQLLOQLATTCDDDTTSLRLDCLYLEEKERLKEEWTLFKEQKNFER
polyA_signal	RSFTEAAIRLGRKAFEEERASWVKQOFLNMTNFHONSENKVLFSAFSGSSDPDLN IVHSRPROKHLHSHVANGVPACTSKLTSLPASPTSDPROTHSCVSHSSISVLNITP EESKPEVARESTDQKNSVQSRPSRSGCGYSCSSAFSAHGDRDDL P"
polyA_site	3168. .3173 /note="putative" 3185 /note="putative"
ORIGIN	
Query Match	99.4%; Score 2676; DB 6; Length 3185;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 2689; Conservative	0; Mismatches 0; Indels 3; Gaps 1;
Qy	1 CGTAGGAGCTGACAGGAGCTGTGTAAGCGTGCAGCAGCTAGCCGCCCTCTCAGGTAT 60
Db	310 CGTAGGAGCTGACAGGAGCTGTGTAAGCGTGCAGCAGCTAGCCGCCCTCTCAGGTAT 369
Qy	61 CTGGGCTCTGAACTTGTCTATGGAGATTGGATGCTGTGACAGATCCAGTTCTGTGTAC 120
Db	370 CTGGGCTCTGAACTTGTCTATGGAGATTGGATGCTGTGACAGATCCAGTTCTGTGTAC 429
Qy	121 AGAAACAAAATCTCTCAATATACCTCAGAAACAAAGATGTCTCGTCCAGTTGTA 180
Db	430 AGAAACAAAATCTCTCAATATACCTCAGAAACAAAGATGTCTCGTCCAGTTGTA 489
Qy	181 CTCCAGCAAGTTCTGTGCTCTTCACTACCTTTATCCAAAACGTCATGCTGTGTTTCGG 240
Db	490 CTCCAGCAAGTTCTGTGCTCTTCACTACCTTTATCCAAAACGTCATGCTGTGTTTCGG 549
Qy	241 TGTCTTTCGACAGGAGAGAACTTGAACAAAGATTTCTTCTATCGAGAGCTGAC 300
Db	550 TGTCTTTCGACAGGAGAGAACTTGAACAAAGATTTCTTCTATCGAGAGCTGAC 609
Qy	301 CACCTTCGGTTTCTCTTCTTGTATGAAGATCCAAAATTAAGAGGCAAGAGAGAAAT 360
Db	610 CACCTTCGGTTTCTCTTCTTGTATGAAGATCCAAAATTAAGAGGCAAGAGAGAAAT 669
Qy	361 AATATAGTCGCTTCTGAACTGTATGAACGAGCTCTCGTGTCTTCAAGGAGAGAACCT 420
Db	670 AATATAGTCGCTTCTGAACTGTATGAACGAGCTCTCGTGTCTTCAAGGAGAGAACCT 729
Qy	421 GCTGGCCAGGAGAGCGTGGAGACACAACTTGAAGCTGGGAGTGCATGAGACCACT 480
Db	730 GCTGGCCAGGAGAGCGTGGAGACACAACTTGAAGCTGGGAGTGCATGAGACCACT 789
Qy	481 GCAGAGCTGCTACGCCAACTTAAAGAGAGCTTGGAAACCTCCAGCGGGAGATGTCGG 540
Db	790 GCAGAGCTGCTACGCCAACTTAAAGAGAGCTTGGAAACCTCCAGCGGGAGATGTCGG 849
Qy	541 GCTTCAAGAGAGAGACAGGAGCTGCAGTCAAGAACAGAGAGTTTGCATCAGCTCTCGAA 600
Db	850 GCTTCAAGAGAGAGACAGGAGCTGCAGTCAAGAACAGAGAGTTTGCATCAGCTCTCGAA 909
Qy	601 GAATGAGAAAGATGAGGTAACAAAATTAACAAAATATCATAGCCAGCGGGCTACTCAGTA 660
Db	910 GAATGAGAAAGATGAGGTAACAAAATTAACAAAATATCATAGCCAGCGGGCTACTCAGTA 969
Qy	661 TAATCATGATGTGAAGAGGAGGCGTGAATATAAAGCTAAAGAGGCGCTGCATCA 720
Db	970 TAATCATGATGTGAAGAGGAGGCGTGAATATAAAGCTAAAGAGGCGCTGCATCA 1029
Qy	721 GCTCGTTTATGAAACAAAGAGATAAAAACATAGCCATGGATGTTTTTAAATTTATGTGGGTG 780
Db	1030 GCTCGTTTATGAAACAAAGAGATAAAAACATAGCCATGGATGTTTTTAAATTTATGTGGGTG 1089
Qy	781 AGCTGATGGCAACAGAGCTCATGAGGAGCTGACAAACAGAACCGAGGAATGAAGATGA 840
Db	1090 AGCTGATGGCAACAGAGCTCATGAGGAGCTGACAAACAGAACCGAGGAATGAAGATGA 1149
Qy	841 GATGTACAAAATCTCTGTTGAATGATTAAGTACCGCCAGAGAGCAGATCTCTGATGGAGAA 900
Db	1150 GATGTACAAAATCTCTGTTGAATGATTAAGTACCGCCAGAGAGCAGATCTCTGATGGAGAA 1209

QY	901	CGCGGAGCTGAAGAAGTCTCCAGCAGATGAGAGGAGATGATCTCTCTCTCTCTCC	960
Db	1210	CGCGGAGCTGAAGAAGTCTCCAGCAGATGAGAGGAGATGATCTCTCTCTCTCTCC	1269
QY	961	TCAGAGAAAGAGCCAGGGAAGAGCAGAGGACGGCAGGCACTGTGCTATCTCCGA	1020
Db	1270	TCAGAGAAAGAGCCAGGGAAGAGCAGAGGACGGCAGGCACTGTGCTATCTCCGA	1329
QY	1021	TATAGAGATGATCTTGGGGAACGTGAGCAGAGACAGCTGTGGGCGCTTTCGTGACAC	1080
Db	1330	TATAGAGATGATCTTGGGGAACGTGAGCAGAGACAGCTGTGGGCGCTTTCGTGACAC	1389
QY	1081	TGTGAGAGAGCAGCTGACAAAAGCATCAGGAAACAGTGGGAAATTTGAAAAGTCATGT	1140
Db	1390	TGTGAGAGAGCAGCTGACAAAAGCATCAGGAAACAGTGGGAAATTTGAAAAGTCATGT	1449
QY	1141	AGAAAACCTCGATAACCAAGCTTCGAAGGTACACTCAGAGGCGCTTAATGAGGAGACGT	1200
Db	1450	AGAAAACCTCGATAACCAAGCTTCGAAGGTACACTCAGAGGCGCTTAATGAGGAGACGT	1509
QY	1201	GATCTCAGCAGAACCATGAGCAAGACTGAGAAACTGGAGCTGGAGATTGACGGTG	1260
Db	1510	CATCTCAGCAGAACCATGAGCAAGACTGAGAAACTGGAGCTGGAGATTGACGGTG	1569
QY	1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGTGTGA	1320
Db	1570	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGTGTGA	1626
QY	1321	TGATGACACCACTCACTGTTCCGAGACTGTTACTTGTGGAAGAAAAGGAACGCTTAA	1380
Db	1627	TGATGACACCACTCACTGTTCCGAGACTGTTACTTGTGGAAGAAAAGGAACGCTTAA	1686
QY	1381	AGAGAGTGGACCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGCGAAGCTT	1440
Db	1687	AGAGAGTGGACCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGCGAAGCTT	1746
QY	1441	TACAGAACTGCCATTCGATTTGGGCTTGGAGAGAAAGCGTTTGAAGAGCGAGCCAG	1500
Db	1747	TACAGAACTGCCATTCGATTTGGGCTTGGAGAGAAAGCGTTTGAAGAGCGAGCCAG	1806
QY	1501	CTGGTGAAGCAGCAGTGTAAAAACATGACGAACCTTTGACCAACCAAGAACTCAGAAAATGT	1560
Db	1807	CTGGTGAAGCAGCAGTGTAAAAACATGACGAACCTTTGACCAACCAAGAACTCAGAAAATGT	1866
QY	1561	GAACTTTTCACTGCTCTCAGGAAGTTCTGATCCAGCAATCTTATAGTCCACTCAG	1620
Db	1867	GAACTTTTCACTGCTCTCAGGAAGTTCTGATCCAGCAATCTTATAGTCCACTCAG	1926
QY	1621	GCCACGGCAAGAGCTACAGTGTGGCTAATGGGTGCGAGCTTGCAACATCAAACT	1680
Db	1927	GCCACGGCAAGAGCTACAGTGTGGCTAATGGGTGCGAGCTTGCAACATCAAACT	1986
QY	1681	GACTAAATCTCTCTGCTCCTACCTTCTACTTTCAGACTTTGCGCAGACACATTCATGT	1740
Db	1987	GACTAAATCTCTCTGCTCCTACCTTCTACTTTCAGACTTTGCGCAGACACATTCATGT	2046
QY	1741	GTCGAAACACAGTTCATCAGTGTGCTGAATATAAATCTCTGAAGAAAGTAAACCAAGTGA	1800
Db	2047	GTCGAAACACAGTTCATCAGTGTGCTGAATATAAATCTCTGAAGAAAGTAAACCAAGTGA	2106
QY	1801	GGTTGCAAGAGAAAGCAGGATCAGAACTGAGCGCTGAGTCGAGGCCACCTCGCGGA	1860
Db	2107	GGTTGCAAGAGAAAGCAGGATCAGAACTGAGCGCTGAGTCGAGGCCACCTCGCGGA	2166
QY	1861	GGGTGCTACAGCGGATGCTCTCGGCTTCAGGAGCGCTCACGGGACCGAGATGACTT	1920
Db	2167	GGGTGCTACAGCGGATGCTCTCGGCTTCAGGAGCGCTCACGGGACCGAGATGACTT	2226
QY	1921	ACCTTAAATGTGCGGCTGCACTGCTGTTCAGATGTGCGCTAGAGAGATTGACACAGG	1980
Db	2227	ACCTTAAATGTGCGGCTGCACTGCTGTTCAGATGTGCGCTAGAGAGATTGACACAGG	2286
QY	1981	GTGTAGCATAAAGTCAGTCGCTAACTTAAGATGCTCAGAGTTGTTGTTGGACTTCGC	2040

Db	2287	GTGTAGCATAAAGTCAGTCGTAACTTAAGATGCTCAGAGTTGTTGTTGGACTTCGC	2346
QY	2041	TGCTTTCCCCCAAGAGCTGAAATCTAAGCTTACCTTAAAAAGGATGCAAAAGCTTTGGTTGT	2100
Db	2347	TGCTTTCCCCCAAGAGCTGAAATCTAAGCTTACCTTAAAAAGGATGCAAAAGCTTTGGTTGT	2406
QY	2101	GTGTTAGTAACAGAAAGCCCTGGCTCTGTGACTGCAGGAATGCATGGCGTTTGGATGGAA	2160
Db	2407	GTGTTAGTAACAGAAAGCCCTGGCTCTGTGACTGCAGGAATGCATGGCGTTTGGATGGAA	2466
QY	2161	ACAGAGCGCTCGAATGATTCGCTCGCCAGGTACCGAGAAGAGCACTTTTAGGGAAGTGT	2220
Db	2467	ACAGAGCGCTCGAATGATTCGCTCGCCAGGTACCGAGAAGAGCACTTTTAGGGAAGTGT	2526
QY	2221	TCCTGTAACACATTAATAATTCGTCCTCAAGTGTGGTTGGCATTTGGAAGTGTAGCCTTACT	2280
Db	2527	TCCTGTAACACATTAATAATTCGTCCTCAAGTGTGGTTGGCATTTGGAAGTGTAGCCTTACT	2586
QY	2281	TGAATGTATCTGTAGATTTTAAACAAAGCAGGTTCTATATTTATTTAGTGTGTAT	2340
Db	2587	TGAATGTATCTGTAGATTTTAAACAAAGCAGGTTCTATATTTATTTAGTGTGTAT	2646
QY	2341	TTTGGGATTACCTCTTTTCATATGTTTGTGCTGTACATAAATACATGACTATGTAA	2400
Db	2647	TTTGGGATTACCTCTTTTCATATGTTTGTGCTGTACATAAATACATGACTATGTAA	2706
QY	2401	GAGGCTTTAAGCTTTAAACAACTTCACACCATGCTTGTAGTATAGCATTTTCATGCCAATTAA	2460
Db	2707	GAGGCTTTAAGCTTTAAACAACTTCACACCATGCTTGTAGTATAGCATTTTCATGCCAATTAA	2766
QY	2461	AATGTTTTCAGTGGCATGGTGTTCACAGAGTTTAGGACCCTGCCACATGACAGTTAAGA	2520
Db	2767	AATGTTTTCAGTGGCATGGTGTTCACAGAGTTTAGGACCCTGCCACATGACAGTTAAGA	2826
QY	2521	CTTTATTTTAAAGCCTCGGCAATAAAAATTCAAAGCCCTTCATTAAGCTGAGTTTCAAG	2580
Db	2827	CTTTATTTTAAAGCCTCGGCAATAAAAATTCAAAGCCCTTCATTAAGCTGAGTTTCAAG	2886
QY	2581	ATAACTAGAACTACTTAACGTTTACATTTTGTAGATTTTAAAGCATTTGATTTTATTTAT	2640
Db	2887	ATAACTAGAACTACTTAACGTTTACATTTTGTAGATTTTAAAGCATTTGATTTTATTTAT	2946
QY	2641	ATATGTGAATTTATAATTTCTAAGAGAAATTTGATTTATGAGTAATGGGG	2692
Db	2947	ATATGTGAATTTATAATTTCTAAGAGAAATTTGATTTATGAGTAATGGGG	2998
RESULT 2			
AK031356			
LOCUS			
DEFINITION			
Mus musculus 13 days embryo male testis cDNA, RIKEN full-length			
enriched library, clone:603041112 product:hypothetical protein,			
full insert sequence.			
ACCESSION			
AK031356			
VERSION			
AK031356.1 GI:26327238			
KEYWORDS			
HTC; CAP trapper.			
SOURCE			
Mus musculus (house mouse)			
ORGANISM			
Mus musculus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;			
Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE			
1			
Carninci, P. and Hayashizaki, Y.			
High-efficiency full-length cDNA cloning			
Meth. Enzymol. 303, 19-44 (1999)			
REFERENCE			
2			
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,			
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
Normalization and subtraction of cap-trapper-selected cDNAs to			
prepare full-length cDNA libraries for rapid discovery of new genes			
Genome Res. 10 (10), 1617-1630 (2000)			

PUBMED REFERENCE AUTHORS	11042159 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 11076861
TITLE JOURNAL PUBMED REFERENCE AUTHORS	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001) 5 The FANTOM Consortium, the RIKEN Genome Exploration Research Group Phase I and II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)
TITLE JOURNAL PUBMED REFERENCE AUTHORS	6 RIKEN Genome Exploration Research Group, Genome Science Group (Genome Network Core Team) and the FANTOM Consortium. Antisense Transcription in the Mammalian Transcriptome Science 309, 1564-1566 (2005) 7 The FANTOM Consortium, Riken Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). The Transcriptional Landscape of the Mammalian Genome Science 309, 1559-1563 (2005) 8 (bases 1 to 3413) Adachi, J., Aizawa, K., Akimura, T., Akakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayate, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kigawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, R., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akai, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
TITLE JOURNAL	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers 1. 3413 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM,DB:6030411112" /db_xref="taxon:10090" /clone="6030411112" /sex="male" /tissue_type="testis"
FEATURES source	1. 3413 /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="13 days embryo" 264. .2108 /notes="unnamed protein product; hypothetical protein (evidence: ProCrest) putative" /codon_start=1 /protein_id="BAC27363.1" /db_xref="GI:26327239" /translations="MGDMVTVDVLCETNKNLSQYSETKMPSSLYSQVLCSSVP LSKNVHGVFCFTGENIQSISYLDQELTTFPGPSLYEESKKEAKREINIYAVLNC MNELLVQRNLLAOSVETQNLKGLSDMDHLQSCYAKLKEQLETSSRMKLGQERDR QLOCKNRSLLHLLKNEKQVOKLONIIASRATQYNHDVKKREYNKRLKRLQVLVNR KDKNTAMDVLNVYGRADSKRGSRWTDKTEARNEDEMYKILLNDYVROKQIIMENAE LKQQLQMKKEMISLSPSKKPRERABDGTVAISDIEDDSGEISRDSVWGLSCDT VREQLTNKQWRLILKSHVQKLNQASKVHSEGLNEEDVISRODHEQETEKLELEIE RKEMIKAQQLQALQATTCDDDTTSLRLDCYLLERKELKEEMTFLFKQKQKFRER RSFTEAAIRLGLERKAFEBERASWQQLFMNTMFDHNSENKVLFSFGSSDPDL IVHSRPROKLLHSAVGPACTSKLTKSLPASPTSDFRQTHSCVSEHSISVLNITP ESKPSEVARESTDQKWSQSRPSRSGCYSGCSSAFRSAHGDRDDLPL" 3394. .3399 /notes="putative" 3413 /notes="putative" polyA_signal polyA_site ORIGIN Query Match 99.4%; Score 2676; DB 6; Length 3413; Best Local Similarity 99.9%; Pred. No. 0; Matches 2689; Conservative 0; Mismatches 0; Indels 3; Gaps 1; Qy 1 CGTAGAGAGTGCAGGAGCTGTGTAAGCGTCGACGACTGAGCCGCTCCTCAGGTAT 60 Db 185 CGTAGAGAGTGCAGGAGCTGTGTAAGCGTCGACGACTGAGCCGCTCCTCAGGTAT 244 Qy 61 CTTGGCTTGGAACTTCTGCTATGGAGATGGATGACTGTGACAGATCCAGTCTCTGTATC 120 Db 245 CTTGGCTTGGAACTTCTGCTATGGAGATGGATGACTGTGACAGATCCAGTCTCTGTATC 304 Qy 121 AGAAACAAAATCTCTCAATATACCTCAGAAACAAAGATCTCCGTCAGTTTGTGA 180 Db 305 AGAAACAAAATCTCTCAATATACCTCAGAAACAAAGATCTCCGTCAGTTTGTGA 364 Qy 181 CTCCAGCAAGTCTGCTGCTTCTCAGTACCTTTATCAAAACGTCATCGTGTTCGG 240 Db 365 CTCCAGCAAGTCTGCTGCTTCTCAGTACCTTTATCAAAACGTCATCGTGTTCGG 424 Qy 241 TGTCTTCTGCACAGGAGAGAACTTGAACAAAGATTTTCTATCTTGTATCAGAGCTGAC 300 Db 425 TGTCTTCTGCACAGGAGAGAACTTGAACAAAGATTTTCTATCTTGTATCAGAGCTGAC 484 Qy 301 CACCTTGGGTTTCTCTTCTTGTATGAAGATCCAAAGTAAAGAGCAAGAGAGAAAT 360 Db 485 CACCTTGGGTTTCTCTTCTTGTATGAAGATCCAAAGTAAAGAGCAAGAGAGAAAT 544 Qy 361 AAATATAGTCCGCTGCTGAACTGTATGAACAGAGCTGCTGCTTCCAGCGGAAGACCT 420 Db 545 AAATATAGTCCGCTGCTGAACTGTATGAACAGAGCTGCTGCTTCCAGCGGAAGACCT 604 Qy 421 GCTGGCCCGAGAGAGCGTGGAGACACAGAACTTTGAAGCTGGGAGTGCATGACACCACT 480 Db 605 GCTGGCCCGAGAGAGCGTGGAGACACAGAACTTTGAAGCTGGGAGTGCATGACACCACT 664 Qy 481 GCAGAGCTGCTAGCCAAACCTTAAGAGAGAGTTGGAAACGTCAGGGGGAGAGTATCGG 540 Db 665 GCAGAGCTGCTAGCCAAACCTTAAGAGAGAGTTGGAAACGTCAGGGGGAGAGTATCGG 724 Qy 541 GCTTCAAGAGAGAGACAGGCGAGCTGCAGTGCAGAGACAGAGATTTGCATCAGCTCTGAA 600 Db 725 GCTTCAAGAGAGAGACAGGCGAGCTGCAGTGCAGAGACAGAGATTTGCATCAGCTCTGAA 784 Qy 601 GAATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCCGGGCTACTCAGTA 660 Db 785 GAATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCCGGGCTACTCAGTA 844

```
QY 661 TAATCATGATGTGAAGAGGAGCGTGAATATATATATAGCTAAAGGAGCGCTGTCATCA 720
Db |||||
QY 845 TAATCATGATGTGAAGAGGAGGCGTGAATATATATATAGCTAAAGGAGCGCTGTCATCA 904
Db |||||
QY 721 GCTCGTTATGAACAAGAGGATAAAAAATAGCCATGGATGTTTAAATATATGTGGTGG 780
Db |||||
QY 905 GCTCGTTATGAACAAGAGGATAAAAAATAGCCATGGATGTTTAAATATATGTGGTGG 964
Db |||||
QY 781 AGCTGATGCAACAGGAGGCTCATGGAGGACTGACAAAAAGAGGAGGAGGAGGAGGAG 840
Db |||||
QY 965 AGCTGATGCAACAGGAGGCTCATGGAGGACTGACAAAAAGAGGAGGAGGAGGAGGAG 1024
Db |||||
QY 841 GATGTACAAAATTTCTGTTCAATGNTTATGATACCGCCAGAGGAGGAGGAGGAGGAG 900
Db |||||
QY 1025 GATGTACAAAATTTCTGTTGAATGATTATGATACCGCCAGAGGAGGAGGAGGAGGAG 1084
Db |||||
QY 901 CGCGGAGCTGAAGAAGGTCCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTGCTCC 960
Db |||||
QY 1085 CGCGGAGCTGAAGAAGGTCCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTGCTCC 1144
Db |||||
QY 961 TCAGAAGAAGAGCCAGGGAAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
Db |||||
QY 1145 TCAGAAGAAGAGCCAGGGAAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1204
Db |||||
QY 1021 TATAGAGATGACTCTGGGGAAGTCTGAGCAGAGCAGAGGAGGAGGAGGAGGAGGAGG 1080
Db |||||
QY 1205 TATAGAGATGACTCTGGGGAAGTCTGAGCAGAGCAGAGGAGGAGGAGGAGGAGGAGG 1264
Db |||||
QY 1081 TGTGAGAGAGCAGCTGACAAAAGCAGATCAGGAAAAGTGGAGAAATTTGAAAAGTCATGT 1140
Db |||||
QY 1265 TGTGAGAGAGCAGCTGACAAAAGCAGATCAGGAAAAGTGGAGAAATTTGAAAAGTCATGT 1324
Db |||||
QY 1141 AGAAAAAATCTCGATAACCAAGCTTCGAAGGTACACTCAGAGGAGGAGGAGGAGGAGG 1200
Db |||||
QY 1325 AGAAAAAATCTCGATAACCAAGCTTCGAAGGTACACTCAGAGGAGGAGGAGGAGGAGG 1384
Db |||||
QY 1201 CATCTCAGCAGAACCACTGAGCAGAGACTGAGAACTGAGAGCTGGAGATGAGCGGTG 1260
Db |||||
QY 1385 CATCTCAGCAGAACCACTGAGCAGAGACTGAGAACTGAGAGCTGGAGATGAGCGGTG 1444
Db |||||
QY 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGCTGTA 1320
Db |||||
QY 1445 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGCTGTA 1501
Db |||||
QY 1321 TGATGACACCACTCACTCTGTCGAGACTGTTACTTGTCTGGAAGAAAAGGAAAGGCTTAA 1380
Db |||||
QY 1502 TGATGACACCACTCACTCTGTCGAGACTGTTACTTGTCTGGAAGAAAAGGAAAGGCTTAA 1561
Db |||||
QY 1381 AGAAGAGTGGACCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGGCGAAGCTT 1440
Db |||||
QY 1562 AGAAGAGTGGACCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGGCGAAGCTT 1621
Db |||||
QY 1441 TACAGAAGCTGCAATTCGATTGGGCTTGAGAGAAAGGCGTTTGAAGAAGAGCGAGCCAG 1500
Db |||||
QY 1622 TACAGAAGCTGCAATTCGATTGGGCTTGAGAGAAAGGCGTTTGAAGAAGAGCGAGCCAG 1681
Db |||||
QY 1501 CTGGGTAAGCAGCAGTGTATTAACATGACGAACCTTTGACCAACCAAGAACTCAGAAAAATGT 1560
Db |||||
QY 1682 CTGGGTAAGCAGCAGTGTATTAACATGACGAACCTTTGACCAACCAAGAACTCAGAAAAATGT 1741
Db |||||
QY 1561 GAAACTTTTTCAGTGCTCTCTCAGGAAGTTCGATCCAGACAATCTTATAGTCCACTCAG 1620
Db |||||
QY 1742 GAAACTTTTTCAGTGCTCTCTCAGGAAGTTCGATCCAGACAATCTTATAGTCCACTCAG 1801
Db |||||
QY 1621 GCCCGCAAAAGAGCTACACAGTGTGCTTAATGGGGTGCAGCTTGCACATCAAACT 1680
Db |||||
QY 1802 GCCCGCAAAAGAGCTACACAGTGTGCTTAATGGGGTGCAGCTTGCACATCAAACT 1861
Db |||||
QY 1681 GACTAAATCTCTCTCTGCTCCTCACCCTTCTACTTTCAGACTTTGCCAGACACATTCATGTGT 1740
Db |||||
QY 1862 GACTAAATCTCTCTCTGCTCCTCACCCTTCTACTTTCAGACTTTGCCAGACACATTCATGTGT 1921
Db |||||
```

RESULT 3
AK077112
LOCUS
DEFINITION
Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4933406N16 product:HYPOTHETICAL 71.0 KDA PROTEIN
homolog [Mus musculus], full insert sequence.
AK077112
ACCESSION
VERSION AK077112.1 GI:26097206

AK077112 2245 bp mRNA linear HTC 02-SEP-2005
Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4933406N16 product:HYPOTHETICAL 71.0 KDA PROTEIN
homolog [Mus musculus], full insert sequence.

```
QY 1741 GTCTGAACACACAGTTCATCAGTGTGCTGAATATATATATCTCTCTGAAGAAAGTAAACCAAGTGA 1800
Db |||||
QY 1922 GTCTGAACACACAGTTCATCAGTGTGCTGAATATATATCTCTCTGAAGAAAGTAAACCAAGTGA 1981
Db |||||
QY 1801 GGTTCGAAGAGAAAGCACGGATCAGAAGTGGAGCGTGCAGTGCAGGCCAGCTCCGCGGGA 1860
Db |||||
QY 1982 GGTTCGAAGAGAAAGCACGGATCAGAAGTGGAGCGTGCAGTGCAGGCCAGCTCCGCGGGA 2041
Db |||||
QY 1861 GGGGTGCTACAGCGGATGCTCCTCGGCCCTTCAGGAGCGCTCACGGGGACCGAGATGACTTT 1920
Db |||||
QY 2042 GGGGTGCTACAGCGGATGCTCCTCGGCCCTTCAGGAGCGCTCACGGGGACCGAGATGACTTT 2101
Db |||||
QY 1921 ACCTTAAATGTGCGGGCTGCAGTGTCTGCCAGATGTCCGCTAGAGAGGATTCACACAGG 1980
Db |||||
QY 2102 ACCTTAAATGTGCGGGCTGCAGTGTCTGCCAGATGTCCGCTAGAGAGGATTCACACAGG 2161
Db |||||
QY 1981 GTGTAGCATAAAGTCAGTGTCTAACTTAAAGATGTCTAGAGTTGTTTTCGACTTCGC 2040
Db |||||
QY 2162 GTGTAGCATAAAGTCAGTGTCTAACTTAAAGATGTCTAGAGTTGTTTTCGACTTCGC 2221
Db |||||
QY 2041 TGTCTTCCCCCAAGAGCTGAAATGCTAAAGCTACTTAAAAAGGATGCAAGACTTTTGGTTGT 2100
Db |||||
QY 2222 TGTCTTCCCCCAAGAGCTGAAATGCTAAAGCTACTTAAAAAGGATGCAAGACTTTTGGTTGT 2281
Db |||||
QY 2101 GTGTAGTAAACAGAAAGCCCTGGCTCTGTGACTGACAGGAATGCATGGCGTTTGGATGGAA 2160
Db |||||
QY 2282 GTGTAGTAAACAGAAAGCCCTGGCTCTGTGACTGACAGGAATGCATGGCGTTTGGATGGAA 2341
Db |||||
QY 2161 ACAGAAGCGCTGGAAATGATTCGCTCGCCAGTACCGAGAAGAGCACTTTTAGGACTCGT 2220
Db |||||
QY 2342 ACAGAAGCGCTGGAAATGATTCGCTCGCCAGTACCGAGAAGAGCACTTTTAGGACTCGT 2401
Db |||||
QY 2221 TCCTGTAAACATTAATATTCGTCCTCAAGTGTGGTTGGCAATGGAAAGTGTAGCCCTTACT 2280
Db |||||
QY 2402 TCCTGTAAACATTAATATTCGTCCTCAAGTGTGGTTGGCAATGGAAAGTGTAGCCCTTACT 2461
Db |||||
QY 2281 TGAATGTATACGTAGATTTTAAACAAAGCAGGTTCTATATATATATATATATATATATATATAT 2340
Db |||||
QY 2462 TGAATGTATACGTAGATTTTAAACAAAGCAGGTTCTATATATATATATATATATATATATATAT 2521
Db |||||
QY 2341 TTTGGGATTACTCTTTTCATATGTTTGTGCTGTACATAAATATACATGACTATGTTAA 2400
Db |||||
QY 2522 TTTGGGATTACTCTTTTCATATGTTTGTGCTGTACATAAATATACATGACTATGTTAA 2581
Db |||||
QY 2401 GAGGCTTTAAGGTTTAAAAACTTCACACCATGCTTGGATATAGCATTTTATGCCCAATTAA 2460
Db |||||
QY 2582 GAGGCTTTAAGGTTTAAAAACTTCACACCATGCTTGGATATAGCATTTTATGCCCAATTAA 2641
Db |||||
QY 2461 AATGTTTTCAGTGGCATGGTGTTCACAGAGTTAGGACCACCTGCCACATGACAGTTAAGA 2520
Db |||||
QY 2642 AATGTTTTCAGTGGCATGGTGTTCACAGAGTTAGGACCACCTGCCACATGACAGTTAAGA 2701
Db |||||
QY 2521 CTTTATTTTAAAGCCATCTGGGCAATAAAAATTCAAAAGCCCTTCATAGCTGAGTTCAG 2580
Db |||||
QY 2702 CTTTATTTTAAAGCCATCTGGGCAATAAAAATTCAAAAGCCCTTCATAGCTGAGTTCAG 2761
Db |||||
QY 2581 ATAACTAGAACTACTAACGTTTACATTTTTCAGATTTTAAAGCATTTGATTTTATTTTAT 2640
Db |||||
QY 2762 ATAACTAGAACTACTAACGTTTACATTTTTCAGATTTTAAAGCATTTGATTTTATTTAT 2821
Db |||||
QY 2641 ATATGGAATGTTATATAATTTCTTAAGAGAAATATATGATATGAGGTAATGGGG 2692
Db |||||
QY 2822 ATATGGAATGTTATATAATTTCTTAAGAGAAATATGATATGAGGTAATGGGG 2873
Db |||||
```


QY 661 TAATCATGATGTGACAGAGGAGCGGTGAATATATAATAGCTTAAGGACGGCTGCATCA 720
DB |||||
QY 906 TAATCATGATGTGACAGAGGAGCGGTGAATATATAATAGCTTAAGGACGGCTGCATCA 965
DB |||||
QY 721 GCTCGTTATGAACAAGGATAAAAAATAGCCATGGATGTTTAAATTTATGTGGTGG 780
DB |||||
QY 966 GCTCGTTATGAACAAGGATATAAAACATAGCCATGGATGTTTAAATTTATGTGGTGG 1025
DB |||||
QY 781 AGCTGATGCAAAACGAGGCTCATGAGGACTGACAAAACAGAACGACGGAATGAAGATGA 840
DB |||||
QY 1026 AGCTGATGCAAAACGAGGCTCATGAGGACTGACAAAACAGAACGACGGAATGAAGATGA 1085
DB |||||
QY 841 GATGTACAAAATCTGTTCAATGATTATGATACGACGACGACGACGATCTCTCTCTCTCC 900
DB |||||
QY 1086 GATGTACAAAATCTGTTCAATGATTATGATACGACGACGACGACGATCTCTCTCTCTCC 1145
DB |||||
QY 901 CGCGGAGCTGAAGAAAGGCTCTCCAGCAGATGAAGAGGAGATGATCTCTCTCTCTCTCC 960
DB |||||
QY 1146 CGCGGAGCTGAAGAAAGGCTCTCCAGCAGATGAAGAGGAGATGATCTCTCTCTCTCTCC 1205
DB |||||
QY 961 TCAGAAGAAAGCCAGGAAAGAGCAGAGGACGCGCACGCTGTTGCTATCTCCGA 1020
DB |||||
QY 1206 TCAGAAGAAAGCCAGGAAAGAGCAGAGGACGCGCACGCTGTTGCTATCTCCGA 1265
DB |||||
QY 1021 TATAGAAGTACTCTGGGAACTGAGCAGAGACGCTGTGGGCGCTTCTCTGTGACAC 1080
DB |||||
QY 1266 TATAGAAGTACTCTGGGAACTGAGCAGAGACGCTGTGGGCGCTTCTCTGTGACAC 1325
DB |||||
QY 1081 TGTGAGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAGAAATCATGT 1140
DB |||||
QY 1326 TGTGAGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAGAAATCATGT 1385
DB |||||
QY 1141 AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGCGCTTAATAGGAGGACGT 1200
DB |||||
QY 1386 AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGCGCTTAATAGGAGGACGT 1445
DB |||||
QY 1201 CATCTCAGCAGACCAATGAGCAGACAGCTGAGAACTGAGCTGGAGATGAGCGGTG 1260
DB |||||
QY 1446 CATCTCAGCAGACCAATGAGCAGACAGCTGAGAACTGAGCTGGAGATGAGCGGTG 1505
DB |||||
QY 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGGCCACCACTGTGA 1320
DB |||||
QY 1506 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTAA--CAGACGTGGCCACCACTGTGA 1562
DB |||||
QY 1321 TGATGACACCACTCACTGTTCCGAGACTGTTACTTGTGGAAGAAAAGGAAACGCTTAA 1380
DB |||||
QY 1563 TGATGACACCACTCACTGTTCCGAGACTGTTACTTGTGGAAGAAAAGGAAACGCTTAA 1622
DB |||||
QY 1381 AGAAGAGTGGACCTTTTAAAGACCAAAAAGAAATTTTGAAGAGAAAGCGGAAGCTT 1440
DB |||||
QY 1623 AGAAGAGTGGACCTTTTAAAGACCAAAAAGAAATTTTGAAGAGAAAGCGGAAGCTT 1682
DB |||||
QY 1441 TACAGAAGCTGCCATTCGATTTGGGTTGAGAGAAAGCGTTTGAAGAGAGCGAGCCAG 1500
DB |||||
QY 1683 TACAGAAGCTGCCATTCGATTTGGGTTGAGAGAAAGCGTTTGAAGAGAGCGAGCCAG 1742
DB |||||
QY 1501 CTGGGTAAAGCAGCAGTGTAAAAATGACGAACCTTTGACCAACCACTCAGAAAAATGT 1560
DB |||||
QY 1743 CTGGGTAAAGCAGCAGTGTAAAAATGACGAACCTTTGACCAACCACTCAGAAAAATGT 1802
DB |||||
QY 1561 GAAACTTTTTCAGTGGCTTCTCAGGAAGTTCGATCCAGACAACTTTATAGTCCACTCAGC 1620
DB |||||
QY 1803 GAAACTTTTTCAGTGGCTTCTCAGGAAGTTCGATCCAGACAACTTTATAGTCCACTCAGC 1862
DB |||||
QY 1621 GCCCGGCAAAAGAGCTACACAGTGTGCTTAATGGGTGCGAGCTTGCAATCAAACT 1680
DB |||||
QY 1863 GCCCGGCAAAAGAGCTACACAGTGTGCTTAATGGGTGCGAGCTTGCAATCAAACT 1922
DB |||||
QY 1681 GACTAAATCTCTTCTGCTCACTTCTCACTTCTGAGACTTTCCGACAGACATTCATGTGT 1740
DB |||||
QY 1923 GACTAAATCTCTTCTGCTCACTTCTCACTTCTGAGACTTTCCGACAGACATTCATGTGT 1982
DB |||||

QY 1741 GTCTGAACACAGTTCCTCATCAGTGTGCTGAATATATACTCTCTCTGAAGAAAGTAAACCAAGTGA 1800
DB |||||
QY 1801 GGTTCGAAGAGAAAGCAGGATCAGAACTGGAGCGTGCAGTCGAGCGCCAGCTCCGGGA 1860
DB |||||
QY 2043 GTTTCGAAGAGAAAGCAGGATCAGAACTGGAGCGTGCAGTCGAGGCCCGAGCTCCGGGA 2102
DB |||||
QY 1861 GGGGTGTACAGCGGATGCTCTCTCGGCTTCCAGGAGCGCTCACGGGACCGAGATGACTT 1920
DB |||||
QY 2103 GGGGTGTACAGCGGATGCTCTCTCGGCTTCCAGGAGCGCTCACGGGACCGAGATGACTT 2162
DB |||||
QY 1921 ACCTTAAATGTGCGGCTGCGAGTGTGTTCCAGATGTGCGTCCAGAGTTGACACAGG 1980
DB |||||
QY 2163 ACCTTAAATGTGCGGCTGCGAGTGTGTTCCAGATGTGCGTCCAGAGTTGACACAGG 2222
DB |||||
QY 1981 GTGTAGCATATAAGTCAAGTCTGCT 2003
DB |||||
QY 2223 GTGTAGCATATAAGTCAAGTCTGCT 2245
DB |||||
RESULT 4
AK049080
LOCUS
DEFINITION Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length
enriched library, clone:C230097E14 product:hypothetical protein,
full insert sequence.
ACCESSION AK049080
VERSION AK049080.1 GI:26339745
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
10349636
2
3 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Sugahara, Y., Shibata, K.,
Itoh, H., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
11042159
3
4 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsubara, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multichannel sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861
4
5 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
6 The FANTOM Consortium, the RIKEN Genome Exploration Research Group
Phase I and II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6
7 RIKEN Genome Exploration Research Group, Genome Science Group
(Genome Network Core Team) and the FANTOM Consortium.

TITLE	Antisense Transcription in the Mammalian Transcriptome	Db	358	CGTAGGAGAGTGACAGGAGCTGTTGTAAGCGTCGACGACTGAGCGCCTCCTCAGGTAT	417
JOURNAL	Science 309, 1564-1566 (2005)				
REFERENCE	7				
AUTHORS	The FANTOM Consortium, Riken Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). The Transcriptional Landscape of the Mammalian Genome	Qy	61	CTTGGCTCTGGAACCTTCTGATGGAGATTGGATGACTGTGACAGATCAGTCTCTGTGTAC	120
TITLE	The Transcriptional Landscape of the Mammalian Genome	Db	418	CCTGGCTCTGGAACCTTCTGATGGAGATTGGATGACTGTGACAGATCAGTCTCTGTGTAC	477
JOURNAL	Science 309, 1559-1563 (2005)				
REFERENCE	8 (bases 1 to 2371)	Qy	121	AGAAACAAAAATCTCTCTCAATATACCTCAGAAACAAAGATGCTCCGCTCCAGTTGTGA	180
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	Db	478	AGAAACAAAAATCTCTCTCAATATACCTCAGAAACAAAGATGCTCCGCTCCAGTTGTGA	537
TITLE	Direct Submission	Qy	181	CTCCAGCAAGTTCTGTGCTCTTTCAGTACCTTTATCCAAAACCGTGCATGGTGTTCGG	240
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9216)	Db	538	CTCCAGCAAGTTCTGTGCTCTTTCAGTACCTTTATCCAAAACCGTGCATGGTGTTCGG	597
REFERENCE	8 (bases 1 to 2371)	Qy	241	TGCTCTTCGACAGGAGAGCAATTTGAACAAAGATTTTCCTATCTTCATCAGGAGCTGAC	300
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	Db	598	TGCTCTTCGACAGGAGAGCAATTTGAACAAAGATTTTCCTATCTTCATCAGGAGCTGAC	657
TITLE	Direct Submission	Qy	301	CACCTTCGGGTTTCTCTTCCTGATGAAGATCCAAAAGTAAAGAGGCAAGAGAGAAAT	360
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9216)	Db	658	CACCTTCGGGTTTCTCTTCCTGATGAAGATCCAAAAGTAAAGAGGCAAGAGAGAAAT	717
REFERENCE	8 (bases 1 to 2371)	Qy	361	AAATATAGTGCCTGTTCTGAACTGTATGAACGAGCTCTCTGTCTTTCAGCGGAAGAACCT	420
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	Db	718	AAATATAGTGCCTGTTCTGAACTGTATGAACGAGCTCTCTGTCTTTCAGCGGAAGAACCT	777
TITLE	Direct Submission	Qy	421	GCTGGCCCGAGAGAGCGTGGAGACACAGAACTTGAAGCTGGGAGTGCATGGAACCACT	480
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9216)	Db	778	GCTGGCCCGAGAGAGCGTGGAGACACAGAACTTGAAGCTGGGAGTGCATGGAACCACT	837
REFERENCE	8 (bases 1 to 2371)	Qy	481	GCAGAGCTGCTAGCCCAAACTTAAGAGAGCAGTTGGAACCTCCAGCGGGAGATGATCGG	540
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	Db	838	GCAGAGCTGCTAGCCCAAACTTAAGAGAGCAGTTGGAACCTCCAGCGGGAGATGATCGG	897
TITLE	Direct Submission	Qy	541	GCTTCAAGAGAGACAGGAGCTGCGAGTCAAGAAACAGAGAGTTTGCATCAGCTCTCGAA	600
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9216)	Db	898	GCTTCAAGAGAGACAGGAGCTGCGAGTCAAGAAACAGAGAGTTTGCATCAGCTCTCGAA	957
REFERENCE	8 (bases 1 to 2371)	Qy	601	GAATGAGAAAGATGAGTACAAAATTTACAAAATTTACATAGCCAGCGGGCTACTCAGTA	660
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	Db	958	GAATGAGAAAGATGAGTACAAAATTTACAAAATTTACATAGCCAGCGGGCTACTCAGTA	1017
TITLE	Direct Submission	Qy	661	TAATCATGATGTGAAGAGGAGCGTGAATTAATAAGCTAAAGAGGAGCGCTGCATCA	720
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9216)	Db	1018	TAATCATGATGTGAAGAGGAGCGTGAATTAATAAGCTAAAGAGGAGCGCTGCATCA	1077
REFERENCE	8 (bases 1 to 2371)	Qy	721	GCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTTTAAATTTATGTGGGTG	780
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	Db	1078	GCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTTTAAATTTATGTGGGTG	1137
TITLE	Direct Submission	Qy	781	AGCTGATGGCAACGAGGCTCATGGAGGACTGCACAAACAGAACCCAGGAAATGAAGATGA	840
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9216)	Db	1138	AGCTGATGGCAACGAGGCTCATGGAGGACTGCACAAACAGAACCCAGGAAATGAAGATGA	1197
REFERENCE	8 (bases 1 to 2371)	Qy	841	GATGTACAAAATTTCTGTTGAATGATTAAGTACCGCCAGAGCAGATCCTGTGAGAGAA	900
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	Db	1198	GATGTACAAAATTTCTGTTGAATGATTAAGTACCGCCAGAGCAGATCCTGTGAGAGAA	1257
TITLE	Direct Submission	Qy	901	CGCGAGCTGAAGAAGTCTCTCCAGCAGATGAAGAAGAGATGATCTCTCTCTCTCTCC	960
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9216)	Db	1258	CGCGAGCTGAAGAAGTCTCTCCAGCAGATGAAGAAGAGATGATCTCTCTCTCTCTCC	1317
REFERENCE	8 (bases 1 to 2371)	Qy	961	TCAGAAAG	1020
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saichou, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	Db	1318	TCAGAAAG	1377
TITLE	Direct Submission	Qy	1021	TATAGAGATGACTCTCGGAACTGAGCAGAGACAGAGTGTGGGCGCTTCTCTGTGACAC	1080
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9216)	Db	1378	TATAGAGATGACTCTCGGAACTGAGCAGAGACAGAGTGTGGGCGCTTCTCTGTGACAC	1437
REFERENCE	8 (bases 1 to 2371)	Qy	1081	TGTGAG	1140

ORIGIN

Query Match 73.3%; Score 1973; DB 6; Length 2371;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 2005; Conservative 0; Mismatches 0; Indels 12; Gaps 2;
 Qy 1 CGTAGGAGAGTACAGGAGCTGTTGTAAGCGTCGACGACTGAGCGGCTCTCAGGTAT 60

Db 1438 TGTGAGAGAGCAGTGCACAAACAGCATCAGGAAACAGTGGAGATTTTGAAGAAGTCATGT 1497
Qy 1141 AGAAAACTCGATAACCAAGCTTCGAGGTACACTCAGAGGGCCCTTAATGAGGAGACCT 1200
Db 1498 AGAAAACTCGATAACCAAGCTTCGAGGTACACTCAGAGGGCCCTTAATGAGGAGACCT 1557
Qy 1201 CATCTCACGACAAAGACCATGACAAAGAGACTCAGAAACTGAGAGCTGGAGATTGACGGTG 1260
Db 1558 CATCTCACGACAAAGACCATGACAAAGAGACTGAGAACTGGAGCTGGAGATTGACGGTG 1617
Qy 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACCTGTGA 1320
Db 1618 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTAA--CAGCAGCTGGCCACCACCTGTGA 1674
Qy 1321 TGATCAGACACCCTCAGTTCGAGAGCTGTTACTTCTGCGAGAGAAAGGAAAGCGCTTAA 1380
Db 1675 TGATCAGACACCCTCAGTTCGAGAGCTGTTACTTGTGCGAGAGAAAGGAAAGCGCTTAA 1734
Qy 1381 AGAAGAGTGGACCCCTTTTAAAGAGCAAAAGAAATTTTGAGAGAGAAAGCGGAAGCTT 1440
Db 1735 AGAAGAGTGGACCCCTTTTAAAGAGCAAAAGAAATTTTGAGAGAGAAAGCGGAAGCTT 1794
Qy 1441 TACAGAAGCTGCAATTCGATTGGGCTTGGAGAGAAAGCGCTTTGAAGAGAGCGAGCCAG 1500
Db 1795 TACAGAAGCTGCAATTCGATTGGGCTTGGAGAGAAAGCGCTTTGAAGAGAGCGAGCCAG 1854
Qy 1501 CTGGGTAAGCAGCAGTGTATTAACATGACGAACTTTGACCAACAGAACTCAGAAAAATGT 1560
Db 1855 CTGGGTAAGCAGCAGTGTATTAACATGACGAACTTTGACCAACAGAACTCAGAAAAATGT 1914
Qy 1561 GAACTTTTCAGTGCCCTCTCAGGAA-----GTTCTGATCCAGACAATCTTATAGT 1611
Db 1915 GAACTTTTCAGTGCCCTCTCAGGAAAGTAAACCAAGTTCGATCCAGACAATCTTATAGT 1974
Qy 1612 CCACCTACGGCCACGGCAAAAGAGCTACACAGTGTGGCTAATGGGGTGCCAGCTTGCAC 1671
Db 1975 CCACCTACGGCCACGGCAAAAGAGCTACACAGTGTGGCTAATGGGGTGCCAGCTTGCAC 2034
Qy 1672 ATCAAACTGACTTAATCTCTCTGCTCCTACCTTCTACTTTCAGACTTTCGCCAGACACA 1731
Db 2035 ATCAAACTGACTTAATCTCTCTGCTCCTACCTTCTACTTTCAGACTTTCGCCAGACACA 2094
Qy 1732 TTCTATGTGTCTGAACACAGTTCATCAGTGTGCTGATATTAATCTCTGAGAAAGTAA 1791
Db 2095 TTCTATGTGTCTGAACACAGTTCATCAGTGTGCTGATATTAATCTCTGAGAAAGTAA 2154
Qy 1792 ACCAAGTGAGGTTCGAGAGAAAGCAACGATCAGAAAGTGGAGCGTGCAGTTCGAGGCCAG 1851
Db 2155 ACCAAGTGAGGTTCGAGAGAAAGCAACGATCAGAAAGTGGAGCGTGCAGTTCGAGGCCAG 2214
Qy 1852 CTCGGGAGGGGTGCTACAGCGAGTGTCTCTCGCCCTTCAGGAGCGCTCACGGGACCG 1911
Db 2215 CTCGGGAGGGGTGCTACAGCGAGTGTCTCTCGCCCTTCAGGAGCGCTCACGGGACCG 2274
Qy 1912 AGATGACTTACTTAAATGCGGGCTGCAGTGTCTTCCAGATGTCGCTAGAGAGT 1971
Db 2275 AGATGACTTACTTAAATGCGGGCTGCAGTGTCTTCCAGATGTCGCTAGAGAGT 2334
Qy 1972 TGACACAGGGGTGAGCATAAAGTCAAGTCTGCTTAACTT 2008
Db 2335 TGACACAGGGGTGAGCATAAAGTCAAGTCTGCTTAACTT 2371

RESULT 5
AY412493
LOCUS Mus musculus HCM4575 gene, VIRTUAL TRANSCRIPT, partial sequence, 1845 bp DNA linear GSS 12-DEC-2003
DEFINITION genomic survey sequence.
ACCESSION AY412493
VERSION AY412493.1 GI:39769458
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 1845)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE
AUTHORS 2 (bases 1 to 1845)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submision
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..1845
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1845
/locus_tag="HCM4575"
Query Match 68.1%; Score 1832; DB 14; Length 1845;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1845; Conservative 0; Mismatches 0; Indels 3; Gaps 1;
Qy 80 ATGGAGATTGGATGACTGTGACAGATCCAGTTCGTGTACAGAAAAAATCTCTCT 139
Db 1 ATGGAGATTGGATGACTGTGACAGATCCAGTTCGTGTACAGAAAAAATCTCTCT 60
Qy 140 CAATATACCTCAGAAAAAAGATGTCCTCCGTCACAGTTTGTACTCCACCAAGTTCGTGC 199
Db 61 CAATATACCTCAGAAAAAAGATGTCCTCCGTCACAGTTTGTACTCCACCAAGTTCGTGC 120
Qy 200 TCCTCAGTACCTTTATCCAAAAACGTGATGGTGTTCGCTGCTTCTGCACAGGAGAG 259
Db 121 TCCTCAGTACCTTTATCCAAAAACGTGATGGTGTTCGCTGCTTCTGCACAGGAGAG 180
Qy 260 AACATTGAACAAAGTATTTCTATCTTGTATCAGAGCTGACCACTTCGGGTTCCTTCC 319
Db 181 AACATTGAACAAAGTATTTCTATCTTGTATCAGAGCTGACCACTTCGGGTTCCTTCC 240
Qy 320 TTGTATGAAGAAATCCAAAAAGTAAAGAGGCAAGAGAGATTTAAATATAGTTCGCTG 379
Db 241 TTGTATGAAGAAATCCAAAAAGTAAAGAGGCAAGAGAGATTTAAATATAGTTCGCTG 300
Qy 380 AACTGTATGAACAGAGCTGCTGCTGCTTTCAGCGGAAGAACCTGCTGGCCAGAGAGCGTG 439
Db 301 AACTGTATGAACAGAGCTGCTGCTGCTTTCAGCGGAAGAACCTGCTGGCCAGAGAGCGTG 360
Qy 440 GAGACACAGAACTTGAAGCTGGGCGAGTGACATGACCACTTCGAGAGCTGTACGCCAAA 499
Db 361 GAGACACAGAACTTGAAGCTGGGCGAGTGACATGACCACTTCGAGAGCTGTACGCCAAA 420
Qy 500 CTTAAGGAGCAGTTGGAAAAACGTCCAGCGGGAGATGATCGGGCTTCAAGAGAGAGACAGG 559
Db 421 CTTAAGGAGCAGTTGGAAAAACGTCCAGCGGGAGATGATCGGGCTTCAAGAGAGAGACAGG 480
Qy 560 CAGCTGCAGTGCAGAAACAGGAGTTTGCATCAGCTCCTCTGAAGAAATGAGAAAGATGAGGTA 619
Db 481 CAGCTGCAGTGCAGAAACAGGAGTTTGCATCAGCTCCTCTGAAGAAATGAGAAAGATGAGGTA 540
Qy 620 CAAAAATTACAAAATATCATAGCCCGGGCTTACTCAGTATATATCATGATGTGAAAGAGG 679

Db 541 CAAAAATTACAAATATCATAGCCAGCGGCTACTCAGTATATATCATGTGTGAAGAGG 600
Qy 680 AAGAGCGTGAATATATAGCTAAAGAGCGCTGATCAGCTCGTTATGAACAAG 739
Db 601 AAGAGCGTGAATATATAGCTAAAGAGCGCTGATCAGCTCGTTATGAACAAG 660
Qy 740 GATAAACATAGCATGATGTTTAAATATATGTGGTGCAGCTGATGCAACAGGAGGC 799
Db 661 GATAAACATAGCATGATGTTTAAATATATGTGGTGCAGCTGATGCAACAGGAGGC 720
Qy 800 TCATGGAGGACTGACAAAAACAGAGCCAGGAATGAAGATGAGATGTACAAATTTCTGTTG 859
Db 721 TCATGGAGGACTGACAAAAACAGAGCCAGGAATGAAGATGAGATGTACAAATTTCTGTTG 780
Qy 860 AATGATTATAGTACCGCCAGAACAGATCTGTATGGAGAACCGCGGAGCTGAAGAGGTC 919
Db 781 AATGATTATAGTACCGCCAGAACAGATCTGTATGGAGAACCGCGGAGCTGAAGAGGTC 840
Qy 920 CTCCAGCAGATGAAGAGGAGATGATCTCTCTGCTCTCTCAGAAAGAGAACCCAGG 979
Db 841 CTCCAGCAGATGAAGAGGAGATGATCTCTCTGCTCTCTCAGAAAGAGAACCCAGG 900
Qy 980 GAAAGAGCAGAGGACGGCACAGGCACTGTTGCTATCTCCGATATAGAAGATGACTCTGGG 1039
Db 901 GAAAGAGCAGAGGACGGCACAGGCAC--TGCTATCTCCGATATAGAAGATGACTCTGGG 957
Qy 1040 GAACTGAGCAGACAGGCTGTGGGGCTTTCCTGTGACACTGTGAGAGAGCCTGACA 1099
Db 958 GAACTGAGCAGACAGGCTGTGGGGCTTTCCTGTGACACTGTGAGAGAGCCTGACA 1017
Qy 1100 AACAGCATCAGGAACAGTGGAGAAATTTGAAAGAGTCACTGAGAAAACTCGATTAACAA 1159
Db 1018 AACAGCATCAGGAACAGTGGAGAAATTTGAAAGAGTCACTGAGAAAACTCGATTAACAA 1077
Qy 1160 GCTTCGAAGGTACACTCAGAGGGCTTAATCAGGAGGACGTCATCTCAGCAACAGACCAT 1219
Db 1078 GCTTCGAAGGTACACTCAGAGGGCTTAATCAGGAGGACGTCATCTCAGCAACAGACCAT 1137
Qy 1220 GAGCAAGAGACTGAGAAACTGAGCTGAGATTTAGCGGCTGTAAGAGATGATCAAGGCT 1279
Db 1138 GAGCAAGAGACTGAGAAACTGAGCTGAGATTTAGCGGCTGTAAGAGAGATGATCAAGGCT 1197
Qy 1280 CAGCAGAGCTCTTACAGCAGCAGCTGSCCACCACGCTGTGATGATGACACCACTCACTG 1339
Db 1198 CAGCAGAGCTCTTACAGCAGCAGCTGCCACACGCTGTGATGATGACACCACTCACTG 1257
Qy 1340 TTGCGAGACTGTTACTTGTCTGGAAGAAAGAAAGCGCCTTAAAGAGAGTGACCCCTTTT 1399
Db 1258 TTGCGAGACTGTTACTTGTCTGGAAGAAAGAAAGCGCCTTAAAGAGAGTGACCCCTTTT 1317
Qy 1400 AAAGAGCAAAAAAGAAATTTTGAAGAGAAAGGCGAAGCTTTACAGAGCTGCCATTGGA 1459
Db 1318 AAAGAGCAAAAAAGAAATTTTGAAGAGAAAGGCGAAGCTTTTACAGAGCTGCCATTGGA 1377
Qy 1460 TTGGGGTTGGAGAGAAAGCGCTTTGAAGAGAGCGGACGCTGGTAAAGCAGCAGTTT 1519
Db 1378 TTGGGGTTGGAGAGAAAGCGCTTTGAAGAGAGCGGACGCTGGTAAAGCAGCAGTTT 1437
Qy 1520 TTAACATGACGAACCTTTGACCACAGAACTCAGAAAAATGTGAAACTTTTCAGTGCCTTC 1579
Db 1438 TTAACATGACGAACCTTTGACCACAGAACTCAGAAAAATGTGAAACTTTTCAGTGCCTTC 1497
Qy 1580 TCAGGAAGTTCTGATCCAGACAATCTTATAGTCACTCAGCGCCACGGCAAGAGGCTA 1639
Db 1498 TCAGGAAGTTCTGATCCAGACAATCTTATAGTCCACTCAGCGCCACGGCAAGAGGCTA 1557
Qy 1640 CACAGTGGGCTTAATGGGGTCCAGCTTGACATCAAAACTGACTAAATCTCTTCTCTGCC 1699
Db 1558 CACAGTGGGCTTAATGGGGTCCAGCTTGACATCAAAACTGACTAAATCTCTTCTCTGCC 1617
Qy 1700 TCACCTTCTACTTCAGACTTTCCGACAGACACATTGATGTGTCTGACACAGTTCCATC 1759
Db 1618 TCACCTTCTACTTCAGACTTTCCGACAGACACATTGATGTGTCTGACACAGTTCCATC 1677

Qy 1760 AGTGTCTGAATAACTCTCTGAAGAAAGTAAACCAAGTCAGGTTGCAAGAGAAAGCACG 1819
Db 1678 AGTGTCTGAATAACTCTCTGAAGAAAGTAAACCAAGTCAGGTTGCAAGAGAAAGCACG 1737
Qy 1820 GATCAGAAAGTGGAGCGTGCAGTCGAGGCCAGCTCGCGGAGGGGTCTACAGCGGATGC 1879
Db 1738 GATCAGAAAGTGGAGCGTGCAGTCGAGGCCAGCTCGCGGAGGGGTCTACAGCGGATGC 1797
Qy 1880 TCCTCGGCTTTCAGGAGCGCTCAGCGGACCGAGATGACTTACCTTAA 1927
Db 1798 TCCTCGGCTTTCAGGAGCGCTCAGCGGACCGAGATGACTTACCTTAA 1845

RESULT 6
BC035580
LOCUS

DEFINITION Homo sapiens synovial sarcoma, X breakpoint 2 interacting protein,
mRNA (cDNA clone IMAGE:5575710).

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BC035580.1 GI:22028261
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homniidae; Homo;
1 (bases 1 to 3154)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.P., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Scapleton,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullan,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

REFERENCE
AUTHORS

Mammalian Gene Collection Program Team
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 3154)
NTH MGC Project
Direct Submission
Submitted (31-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-x@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nhgri.nih.gov
Akhtar N., Ayele, K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brookes,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,F., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,

CONSRMT
TITLE

human and mouse cDNA sequences

JOURNAL
PUBMED
REFERENCE
AUTHORS
CONSRMT
JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 3154)

REMARK
COMMENT

NTH MGC Project
Direct Submission
Submitted (31-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NTH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-x@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nhgri.nih.gov
Akhtar N., Ayele, K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brookes,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,F., Legaspi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,

QY 181 CTCCAGCAAGTTCTGTGCTCTTCACTACCTTTATCCAAAACGGTGCATGGTGTTCGG 240
DB 368 CTCACAGCAAGTGTCTATGTTCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTCAG 427
QY 241 TGTCTTCTGCACAGGAGAAACAATTGAACAAAAGTATTTCTCTATCTTTGATCAGGAGCTGAC 300
DB 428 TGCCTTCTGCACAGAGATAAATTGAACAGAGTATCTCATATCTTGTATCAGGAATTGAC 487
QY 301 CACCTTCGGGTTTCTTCTTGTATGAAGATCCAAAAGTAAAGAGGCAAGAGAAATT 360
DB 498 TACTTTTGGTTTCTCTATTATATGAAGAAATCCAAAGGTAAGAGACAAGAGAGAGTT 547
QY 361 AAATATAGTCGTGTTCTTGAACCTGTATGAACGAGTGTCTGCTTTCAGCGAAGAACT 420
DB 548 AAATATAGTAGTGTACTAAATTGTATGAGCTGTCTTGTCTTTCAGCGGAAGAACT 607
QY 421 GCTGCCAGGAGAGCGTGGAGACACAGAACTTGAAGCTGGGAGTGCATGAGCACT 480
DB 608 TCTAGCTCAGGAAAATGTGGAGACACAGAAATTTGAAGCTGGGAAGTATGACCACT 667
QY 481 GCAGAGCTGTACGCCAAACTTTAAGGAGAGTTGGAACAGTCCAGCGGGAGATGATCGG 540
DB 668 ACAGAGCTGTACTCAAAACTTTAAGGAACAACCTGGAAAACCTCCAGGAGGGAATGATTGG 727
QY 541 GCTTCAAGAGAGACAGCGAGCTGCAGTGCAGAACAGAGGATTTGCACTAGCTCTGAA 600
DB 728 GCTTCAGGAAGAGACAGACAGTTCACAAATGTAAAGAACAGGAATTTGCACTACGTACTAAA 787
QY 601 GAATGAGAAAGATGAGGTACAAAAATACAAAATATCATAGCCAGCCGGCTACTCAGTA 660
DB 788 GAATGAGAAAGATGAGGTGCAAAAATACAAAATATCATTCGAAGTCCAGCTACTCAGTA 847
QY 661 TAATCATGATGTGAAGAGAGGAGCGTGAATATATAAGCTTAAAGGAGCGCTGCATCA 720
DB 848 TAATCATGATGTGAAGAGAAAGAGCGTGAATATATAAATGAAGGAACCGTACATCA 907
QY 721 GCTGTTATGAACAAGAGAGTAAAAACATAGCCATGATGTTTAAATTTATGTGGTGC 780
DB 908 ACTGTTATGAACAAGAAAGATAGAAATAGCTATGGACATTTTGAATTTATGTGGGAG 967
QY 781 AGCTGATGGCAAAACGAGGCTCATGGAGACTGCACAAAACAGAAAGCCAGGAATGAAGATGA 840
DB 968 AGCTGATGGAAAAGAGGCTCTCGAGGACTGGTAAACTGAAAGCCAGGATGAAGATGA 1027
QY 841 GATGTACAAAATTTCTGTTGAATGATTATGATACCGCCAGAGACAGATCTCTGATGGAGAA 900
DB 1028 AATGTATAAAAATTTCTTTGAATGATTATGAATATCGTCAGAAAACAAATCTTAATGGAAA 1087
QY 901 CGCGAGCTGAAGAGGTCCTCCAGCAGATGAAAGAGAGATGATCTCTCTCTCTCTCC 960
DB 1088 TGCAGAACTTGAAGAGGTTCTTCAACAAATGAAAAGGAAATGATTTCTCTCTCTCTCC 1147
QY 961 TCAGAAGAAGAGCCAGGAAAGAGCAGAGACGGCACAGGCACTGTTGCTATCTCGA 1020
DB 1148 CCAAAAGAAAGAACTTAGAGAAAGTAGATGATAGTACAGGAACTGT---TATTTCCGA 1204
QY 1021 TATGAAGATGACTCTGGGGAACTGAGCAGAGACAGCGTGTGGGCGCTTCTCTGTGACAC 1080
DB 1205 TGTGAAGAGATGCGGGGAACTAAAGCAGAGAGATGATGTGGGACCTTCTCTGTGAAAC 1264
QY 1081 TGTGAGAGCAGCTGACAAAACAGCATCAGGAACAGTGGAGAAATTTTGAAGACTCATGT 1140
DB 1265 TGTGAGAGCAGCTTACAAACAGCATCAGAAAACAGTGGAGAAATTTTGAAGAGTCAATGT 1324
QY 1141 AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGCGCTTAATGAGGAGGAGCT 1200
DB 1325 AGAAAAGCTTGAATAACCAAGTTTCAAGGTTACACTGGAAGGTTTAAATGATGAAGATGT 1384
QY 1201 CATCTCAGCAAGACCATGAGCAAGAGACTGAGAAACTGGAAGCTGGAGATTGAGCGGTG 1260
DB 1385 AATCTCAGCAAGACCATGAACAAGAACTGAAAAACTCGAGTTAGAAATTCAGCAGTG 1444
QY 1261 TAAAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCAGTGTGA 1320

DB 1445 TAAAGAAATGATTAATAAATCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTCATATGA 1504
QY 1321 TGATGACACCACTCCTCACTGTTGCGAGACTGTTTACTTGTGGAGAAAAGGAACGCTTAA 1380
DB 1505 TGATGATACCACTTCACTATTACGAGACTGTTTATTTGTTGGNAGAAAAGGAACGCTCAA 1564
QY 1381 AGAAGATGGACCCCTTTTAAAGAGCAAAAAAAGAAATTTTGAGAGAGAAAAGCGAAGCTT 1440
DB 1565 AGAAGAAATGGTCCCTTTTAAAGAGCAGAAAAGAAATTTTGAGAGGGAGAGACGAAGCTT 1624
QY 1441 TACAGAACTGCCATTCGATTGGGTTTGGAGAGAAAGGGTTTGAAGAGAGCGGACCCAG 1500
DB 1625 TACAGAGCCGCTATTTCGCCTGGGATTGGAGAGAAAAGGCATTTGAAGAGAAAGAGCCAG 1684
QY 1501 CTGGTAAGCAGCAGAGTTTTAAACATGACGAACCTTTCACCAACAGAACTCAGAAAAATGT 1560
DB 1685 TTGGTTAAAGCAGCAGTTCCTAATATGACTTACCTTTGACCCAGAACTCAGAAAAATGT 1744
QY 1561 GAAACTTTTTCAGTGCCTTCTCAGGAAGTTCTGATCCAGACAATTTATAGTCCACTCAG 1620
DB 1745 GAAACTTTTTCAGTGCCTTCTCAGGAAGTTCTGATTGGGACAATCTTATAGTCACTCGAG 1804
QY 1621 GCCAGGCAAAAGAGCTACACAGTGTGGCTAATGGGGTGCAGCTTGCACATCAAACT 1680
DB 1805 GCAGCCGCAAAAGAGCCCTCACAGTGTCTAATGGGTCTCCAGTTCATGTCTCTAACT 1864
QY 1681 GACTAAATCTCTCTCGCTCACCCTTCTACACTTTCAGACTTTCCAGACACACATTCATGTGT 1740
DB 1865 TACTAAATCTCTCTCGCTCACCCTTCCACTTTCAGACTTTTCCAGACACAGTTCCTCGAT 1924
QY 1741 GTCTGAACACAGTTCCTCACTGCTGCTGAATATATACTCTGAAAGAAAGTAAACCAAGTGA 1800
DB 1925 ATCTGAACATAGTTCAATCAATGACTGAATATAAATGCTGAAGAAATTAACCAAAATCA 1984
QY 1801 GGTTCGAAGAGAAAGCAGGATCAGAGTGCAGCGTGCAGTGCAGGCCCCAGCTCGCGGA 1860
DB 1985 GGTTCGAGGAGAAAGTACAAATCAAAATGAGGTGTGGCGTCAAGACCTGGATCAACGA 2044
QY 1861 GGGGTGCTACAGCGGATGCTCCTCGGCTTCAGAGCGCTCACGGGGACCCAGAGTACTT 1920
DB 2045 AGTTTGTCTATAGTGTGCTCTTCTGAGCTACAAATCTCATGTAGAAAAGATGACTT 2104
QY 1921 ACCTTAAATGTGCGGGTGCAGTGTGTTCCAGATGTGCGCTAGAGG-----A 1969
DB 2105 ACCTTAGACATGTGGACTGGAAATTTTTCATTAATGTGTTTCATCAAGTTTCACATCTAA 2164
QY 1970 GTTCACACAGGCTAGCATAAAGTCACTGCTC-----TAACTTAAGATGCTCAGAGTTG 2024
DB 2165 GTTGAACACAGGCTGTGTATTAAGTCACTTCTTAATAAAGTGGTCTGAGTTG 2224
QY 2025 TTTGTTTGGACTTCGCTGCTTCCCCCAAGAGCTGAAATGCT--AAGCTACTTAAAGGA 2083
DB 2225 TTTGTTTGGACTTCGCTGCTTCCCCCAAGAGTTGAATCTTAAATCTATTTAAAGGA 2284
QY 2084 TGAAGAGCTTTGG-----TTGTGTGTTAGTAAACAGAACCCCTGCTCTGTGACTGCAAGGA 2139
DB 2285 TATAAAGCTTTGGATATGATTTTGTAGTAACAGAGCATCTGCTTCTGTAATAAGGA 2344
QY 2140 ATGCATGCG--GTTTGGATGGAACAGAGCGCTGGAATGA--TTGCTCTCGCCAGGTACCG 2196
DB 2345 ATGTATAGATGTTTGGATGGAACAAAAGCACTAGACTGAGTTCCTCTTATAGGTATTA 2404
QY 2197 AGAAGAGCACTTTTAGGCACTGGTTCTGTAAACATTTAAATATTCGTGCCAAGTGTGTT 2256
DB 2405 AATAAGCACTTTTAGAAACTGATTAATGTAAATGTTTAAATTTTGTCTCAATATAGTT 2464
QY 2257 GGCATTGGAAGTGTAGCCTTTACTTGAATGTATCTGTAGATTTTAAACAAAGCAGGTT 2316
DB 2465 GGCATTGGAAGTGTAGCCTTTACTTGAATGTATCTGTAGATTTTAAACAAAGCAGGTT 2524
QY 2317 TATATTTATTTAGTGTGTTTGGGATTTACCTCTTTCATATGTTTT----- 2367

Db	2525	TATATTTATGTTTATGTTG-GTTTGAATACTACCTCTTTTCATATGCTTTTAAATAAAGT	2583
Qy	2368	-----GTGCTGTACATAAAATATACATGACTATGTTTAAAGAGGCTTTAAAGGTTT	2415
Db	2584	GAATTTATGATGTTTGTACATAGATACACATGATTATGTTTAAAGAGGCTTTAAGATTT	2643
Qy	2416	AAAAAATCTCACCAATGCTTTGAGTATAGCAATTTTCATGCCAATTAATAATGTTTTCAGTGGC	2475
Db	2644	AAAAAGTTTACA-CAACCATAAGTATAGTATTTTCATGCC-AGTAAATAATTTTTCAGTGGT	2701
Qy	2476	ATGGTGTTTACAGAGG--TTAGACCACTGCCACATGACATGACAGTTTAAAGACTTTTATTTTAAAG	2533
Db	2702	ATTCGTGTTTACAGATGTTATAGGACCAATTGATGCAATTACATTTAAGA-ATTCTCTTTAAT	2760
Qy	2534	CCATCTGGGCAATAAATAATTCAAAGCCCTTCAT-AAGCTGAGTTC---AGATAAAGTACA	2589
Db	2761	ACATCTGGGCAATAAATAATTGAAAGGTATTCCATGAAGCTGAGTCTTTTATAGATAATCAAC	2820
Qy	2590	ACTACTAAGCTTACATTTTGGAGATTTTAAAGCAAT-GTATTTTATTTTATATATATGCT-G	2647
Db	2821	ACTACTAAGCTTACATTTTGGAGATTTTATGACATTTAGATTTTATTTTGTATATGATAG	2880
Qy	2648	AATGTTTATTAATTTCTAAGAGGAATTAATGATATGAGTAATGGG	2692
Db	2881	AATATTATTAATTTTAAAGGACTATTGATGATAGAAGATAGGG	2925
RESULT 8			
CR613292			
LOCUS			
DEFINITION	CR613292 2111 bp mRNA linear HTC 21-JUL-2004		
ACCESSION	full-length cDNA clone CS0DL007Y108 of B cells (Ramos cell line)		
VERSION	Cot 25-normalized of Homo sapiens (human).		
KEYWORDS	CR613292.1 GI:50494099		
SOURCE	HTC; CNSUT_cDNA.		
ORGANISM	Homo sapiens		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;		
JOURNAL	Hominidae; Homo.		
REMARK	1 (bases 1 to 2111)		
	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.		
	Full-length cDNA libraries and normalization		
	Unpublished		
	Contact : Feng Liang Email : fliang@lifetech.com URL :		
	http://fulllength.invitrogen.com/ Invitrogen Corporation 1600		
	Faraday Avenue		
REFERENCE	2 (bases 1 to 2111)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :		
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr		
COMMENT	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime		
	end enriched, double-strand cDNA was digested with Not I and cloned		
	into the Not I and EcoR v sites of the pCMVSPORT 6 vector. Library		
	was normalized. Library was constructed by Life Technologies, a		
	division of Invitrogen.		
FEATURES	Location/Qualifiers		
source	1..2111		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CS0DL007Y108"		
	/tissue_type="B cells (Ramos cell line) Cot 25-normalized"		
	/plasmid="pCMVSPORT_6"		
ORIGIN			
Query Match	52.6%;	Score 1416.6;	DB 6; Length 2111;
Best Local Similarity	83.1%;	Pred. No. 0;	
Matches 1627;	Conservative 0;	Mismatches 329;	Indels 3; Gaps 1;
Qy	1	CGTAGGAGAGTGACAGGAGCTGTTGTAAGCGTCGACAGCACTGAGCCGCTCCTCAGGTAT	60

Db	127	CATAGAAGAGTGACACAGCTAGACTAAATGTTTAACTGCTGAACTAGTTTCCCTCAGGTAT	186
Qy	61	CTGGGCTCTGGAACTTGTCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTAC	120
Db	187	CTGGGCTCTAGAGATTGCTATGGGAGATTGGATGACTGTGTACAGATCCAGGTTCTGTCTTC	246
Qy	121	AGAAACAAAATCTCTCAATATACCTCAGAAAACAAAGATGTCCTCGTCCAGTTTGTGA	180
Db	247	AGAAAGCAAACTATCTCTCAATATACCTCAGAAAACAAAGATGTCCTCCATCAAGTTTATA	306
Qy	181	CTCCAGCAAGTTCTGTGCTCTTTCAGTACCTTTTATCCAAAAAGCTGTCATGGTGTTCGG	240
Db	307	CTCACAGCAAGTCTATGTTCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTCAG	366
Qy	241	TGCTCTTCGACAGGAGAGAACATTGAAACAAAGATTTTCTCTATCTTGTATCAGAGCTGAC	300
Db	367	TGCTCTTCGACAGAGATAAATATTGAAACAGAGTATCTCATATCTTGTGATCAGGAATTTGAC	426
Qy	301	CACCTTCGGGTTTCTCTTCTTGAAGTATGAAGAACTCAAAAGTAAAGAGGCAAGAGAGATT	360
Db	427	TACTTTTGGTTTTCTTCTTATATGAAGAACTCAAAAGTAAAGAGCAAAAGAGAGATT	486
Qy	361	AAATATAGTCTGTTCTGAACTGTATGAACGAGTCTGTCGTCCTTCAGCGGAAGAACCT	420
Db	487	AAATATAGTCTGTTCTTAAATTTGATGAATGAGTCTGTTGCTTCAGCGGAAGAACCT	546
Qy	421	GCTGGCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGGGCGAGTGCACATGACCACCT	480
Db	547	TCTAGCTCAGGAAAAATTTGTGGAGACACAGAAATTTGAAGCTGGGAGTATATGGAACCATCT	606
Qy	481	GCAGAGCTGTACGCCAACTTAAGAGAGCAGTTTGGAAACGTCAGGCGGGAGATGATCGG	540
Db	607	ACAGAGCTGCTCTCTCAAACTTTAAGGAACAACTTGGAAAACCTCCAGGAGGAATGATGG	666
Qy	541	GCTTCAAGAGAGACAGCGCAGCTGCAGTGCAGAAACAGAGAGTTTGCATCAGCTCTCGAA	600
Db	667	GCITTCAGGAAAGAGACAGACAGTTACAATGTAAGAACAGGAAATTTGCATCAGCTACTAAA	726
Qy	601	GAATGAGAAAGATGAGGTACAAAAATTTACAAAATATCATAGCCAGCGGGCTACTCAGTA	660
Db	727	GAATGAGAAAGATGAGGTGCAAAAATTTACAAAATATCATTTGCAAGTCTAGTCTCAGTA	786
Qy	661	TAATCATGATGTCAGAGGAGGAGCGTGAATATAAAGCTAAAGAGGCGCTTCGATCA	720
Db	787	TAATCATGATGTCAGAGGAAAGAGCGTGAATATAAAGCTGAAGGAACTGTACATCA	846
Qy	721	GCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTAAATTTATGTGGGTG	780
Db	847	ACTTGTATTGAACAAGAGATAAGAAAAATAGCTATGGACATTTTGAATTTATGTGGGAG	906
Qy	781	AGCTGATGCCAAACGAGGCTCATGGAGGACTGACAAACAGAGAGCCAGGAATGAAGATGA	840
Db	907	AGCTGATGGAAAAGAGGCTCTCTGGAGGACTGTTAAACCTGAAGCCAGGAATGAAGATGA	966
Qy	841	GATGTACAAAATTTCTGTTGAATGATTTATAGTACCGCCAGAGAGCAGATCTCTGATGGAGAA	900
Db	967	AATGTATAAAATTTCTCTTGAATGATTTATGAATATCGTCAGAACAAATCTTATGGAAAA	1026
Qy	901	CGCGGAGCTGAAGAAGGTCTCTCAGCAGATGAAGAAGAGATGATCTCTCTCTGTCTCC	960
Db	1027	TGCAGAACTTAAAGAGGTTCTTCAACAAATGAAAAAGGAAATGATTTCTCTCTCTCTCC	1086
Qy	961	TCAGAGAGAGAGAGCCAGGAAAGAGAGAGGAGCGGCACAGGACCTGTTGCTATCTCCGA	1020
Db	1087	CCAAAAGAGAGAAACCTTAGAGAAAAGAGTAGATAGTACAGGAACTGT---TATTCCGA	1143
Qy	1021	TATAGAAGATGACTCTGGGGAACCTGAGCAGAGACAGCGTGTGGGGCTTTCTGTGACAC	1080
Db	1144	TGTTGAAGAGATGCCGGGAACTAAGCAGAGAGATGATGTGGACCTTTCTGTGAAAC	1203
Qy	1081	TGTGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAGAGTCATGT	1140

Db	1204	TGTGAGAGAGCAGCTTACAAA	CAGCATCAGAAAA	CAGTGGAGATTTTGA	AAAGTCATGT	1263
Qy	1141	AGAAAACTCGATAACCAAGCTT	CGAAGTACACTCAGAGGCGCTT	PAATGAGAGACGT		1200
Db	1264	AGAAAAGCTTGATAACCAAGTTT	CAAGGTACACCTGGAAAGTTT	TAATGATGAAGATGT		1323
Qy	1201	CATCTCAGCAGAACCATGAC	AGAGAGACTGAGAACTGGAGCT	GGAGATTGAGGGTG		1260
Db	1324	AATCTACAGCAAGACCATGA	ACAGAAACTGAAAACTCGAGT	TAGAAATTCAGCAGTG		1383
Qy	1261	TAAAGAGATGATCAAGGCTT	CAGCAGCAGCTCTTACAG	CAGCAGCTGGCCACCACCGTGGA		1320
Db	1384	TAAAGAAATGATTAAACT	CAGACAGCTTTTACAGCAG	CAGCTCGTACTGTCATATGA		1443
Qy	1321	TGATGACACCACTCAGCTG	TCGGAGACTGTATCTTGTG	GGAAGAAAGCAACGCTTAA		1380
Db	1444	TGATGATACCACTTCACTAT	TACGAGACTGTTATTTGT	TGGAGAAAGCAAGCTCTCA		1503
Qy	1381	AGAAGAGTGGACCTTTTAA	GAGCAAAAGAAATTTGAG	AGAGAAAGCGGAGCTT		1440
Db	1504	AGAAGAAATGGTCCCTTTT	TAAAGAGCAGAAAAAGAA	TTTGGAGGGGAGAGCAGGCTT		1563
Qy	1441	TACAGAAGCTGCCATTCGA	TTCGGGCTTGAGAGAAAGG	CGTTTGAAGAGAGCGGACG		1500
Db	1564	TACAGAAGCCGTATTCG	CCCTGGGATTTGGAGAAAGG	CGATTTGAAGAGAAAGAGCCAG		1623
Qy	1501	CTGGGTAAGCAGCAGTGT	TTTAAACATGACGAACTT	TGACCAACCAAGAACTCAG		1560
Db	1624	TTGGTTAAAGCAGCAGT	TTCTAAATATGACTACCTT	TGACCAACCAAGAACTCAG		1683
Qy	1561	GAAACTTTTCAGTGCTCT	CAGGAGTCTCGATCCAG	CAATCTTATAGTCCACTCAG		1620
Db	1684	GAAACTTTTCAGTGCTCT	CAGGAGTCTCGATCCAG	CAATCTTATAGTCCACTCAG		1743
Qy	1621	GCCAGCGCAAGAGACTAC	AGTGTGGCTTAATGGGGT	GCCAGCTTGCAACATCAAACT		1680
Db	1744	GCAGCGCAAGAGACCTC	AGTGTCTAATGGGTCTC	AGTTTGCATGTCTAACT		1803
Qy	1681	GACTAAATCTCTTCTG	CTCACCCTTCTACTTCA	GACTTTGCCAGACACATTC		1740
Db	1804	TACTAAATCTCTTCTG	CTCACCCTTCTACTTCA	GACTTTGCCAGACACATTC		1863
Qy	1741	GTCGAACACAGTTCAT	CGTGTCTGAATATTAAT	CTCTGAAGAAATGAACCA		1800
Db	1864	ATCTGAACATAGTTCA	ATCAATCAATCAAAAT	CGTCAAGAAATTAACCA		1923
Qy	1801	GTTTGAAGAGAAAGCAG	GATCAGAGTGGAGCTG	CAGTGGCGCCAGCTCGCGGA		1860
Db	1924	GTTTGAAGAGAGATGT	ACAAATCAAAATGAGT	GTGGCATCAAGACCTGGAT		1983
Qy	1861	GGGTGTCTACAGCGAT	CTCTCGCTTCCAGGAG	CGCTCACGGGACCGAGAT		1920
Db	1984	AGGTGTCTATAGTGG	ATGCTCTTGGAGCTCA	CAAAATCTCATGTAGAAA		2043
Qy	1921	ACCTTAATGTGCGGCT	GCAGTGTCTCCAGAT	GTG 1959		
Db	2044	ACCTTACATGTGGACT	GGAAATTTTTCATTA	ATGTG 2082		
RESULT 9						
AY412491						
LOCUS	AY412491					
DEFINITION	Homo sapiens HCM4575 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.	1845 bp	DNA	linear	GSS 12-DEC-2003	
ACCESSION	AY412491					
VERSION	AY412491.1					
KEYWORDS	GSS.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 1845)					

AUTHORS	Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.			
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios			
JOURNAL	Science 302 (5652), 1960-1963 (2003)			
PUBMED	14671302			
REFERENCE	2 (bases 1 to 1845)			
AUTHORS	Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA			
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.			
FEATURES	Location/Qualifiers			
source	1..1845			
gene	/organism="Homo sapiens"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:9606"			
	<1..>1845			
	/locus_tag="HCM4575"			
ORIGIN				
Query Match	50.7%;	Score 1363.8;	DB 14;	Length 1845;
Best Local Similarity	84.0%;	Pred. No. 0;		
Matches 1552;	Conservative	0;	Mismatches 292;	Indels 3; Gaps 1;
Qy	80	ATGGAGATTGGATGACTGTGACAGATCCAGTCTTGTTGTACAGAAAAACAAATCTCTCT	139	
Db	1	ATGGAGATTGGATGACTGTTACAGATCCAGTCTGTCTTCAGAAAGCAAAATCTCTCT	60	
Qy	140	CAATATACCTCAGAAACAAAGATGTCCTCGTCCAGTTTGTACTCCAGCAAGTTCTGTGC	199	
Db	61	CAATATACCTCAGAAACAAAGATGTCCTCATCAAGTTTATATCTACAGCAAGTGTCTATGT	120	
Qy	200	TCTTCAGTACCTTTATCCAAAAAGTGCATCGTGTGTTTTCGGTGTCTTCTGCACAGAGAG	259	
Db	121	TCTTCATACCTTTATCGAAAAATGTGCACAGTTTTTTTCAGTGCCTTCTGCACAGAAAT	180	
Qy	260	AACATTGAACAAAGTATTTCTTATCTTGTATCAGGAGCTGACCACTTCGGGTTTCTCTCC	319	
Db	181	AATATTGAACAGAGTATCTCATATCTTGTATCAGAAATTTGACTACTTTTGGTTTTCTTCA	240	
Qy	320	TTGTATGAAGAAATCCAAAGTAAAGAGCAAGAGAGAAATTAATATATAGTCTGTCTGTG	379	
Db	241	TTATATGAAGAAATCCAAAGGTAAGAGACAAAGAGAGAGTTAAATATATAGTCTGTACTA	300	
Qy	380	AACATGTATGAACAGAGCTGCTGCTGCTTTCAGCGGAAGAACCTGCTGGCCACGAGAGCGTG	439	
Db	301	AATTTGTATGATGAGCTGCTTGTGCTTTCAGCGGAAGAACCTTCTAGCTCAGGAAATGTG	360	
Qy	440	GAGACACAGAACTTGAAGCTGGGAGTGAACATGGACCACTTCGACAGCTGCTACGCCAAA	499	
Db	361	GAGACACAGAAATTTGAAGCTGGGAAGTGAATGACCATCTACAGAGCTGCTACTCAAAA	420	
Qy	500	CTTAAGGAGCAGTTGGAAAAGTCCAGGGGGAGATGATCGGGCTTCAAGAGAGACAGG	559	
Db	421	CTTAAGGAGCAACCTGGAAAGCTCCAGGAGGGGAAATGATTGGGCTTTCAGGAAGAGACAG	480	
Qy	560	CAGCTGAGTGAAGAAACAGAGTTTTCATCAGCTCCTCGAAGAAATGAGAAAGATGAGGTA	619	
Db	481	CAGTTACATGTAAGAACAGGAATTTGCATCAGCTACTAAAGAAATGAGAAAGATGAGGTG	540	
Qy	620	CAAAAATTACAAAATATCATAGCCAGCGGGCTACTCAGTATAATCATGTGTGAAGAGG	679	
Db	541	CAAAAATTACAAAATATCATTTGAAGTCGAGCTACTCAGTATATATCATGTATGAAGAGA	600	
Qy	680	AAGGAGCGTGAATATAATAGCTTAAGAGAGCGCTGCATCAGCTGTTATGAACAGAG	739	


```
Db 241 TTATATGAAGATCCAAAGGTAAAGAGACAAAGAGAGAGATTAAATATAGTAGCTACTA 300
QY 380 AACTGTATGAACAGAGCTGCTGCTTCAGCGGAAAGAA CTTGCTGGCCAGAGAGCGGTG 439
Db 301 AATTGTATGAATGAGCTGCTGCTTCAGCGGAAAGAACTTCTAGCTCAGGAAATGTG 360
QY 440 GAGACACAGAACTTGAAGCTGGGAGTGAGATGGAACCTTCAGAGCTGCTACGCCAA 499
Db 361 GAGACACAGAACTTGAAGCTGGGAGTGATATGGACCATCTACAGAGCTGCTACTCAAA 420
QY 500 CTTAAGGAGCAGTTGGAAACGTCCTCAGCGGAGAGATGATCGGCTTCAAGAGAGAGACAGG 559
Db 421 CTTAAGGAAACAACCTGGAACCTCCAGGAGGGAATGATGGCTTCAGGAAAGAGACAGA 480
QY 560 CAGCTGCAGTGAAGAA CAGGAGTTTTCATCAGCTCCTGAAAGATGAGAAAGATGAGTGA 619
Db 481 CAGTTACAAATGTAAGAA CAGGAAATTTGCATCAGCTACTAAAGAAATGAGAAAGATGAGGTG 540
QY 620 CAAAAATTACAAAATATCATAGCCAGCGCGGCTACTCAGTATAATCATGATGTAAGAGG 679
Db 541 CAAAAATTACAAAATATCATTCGAAGTCGAGCTACTCAGTATAATCATGATATGNNNNN 600
QY 680 AAGGAGCGTGAATATAATAAGCTAAAGGAGCGCCTGCATCAGCTCGTTATGAACAAGAG 739
Db 601 AAGAGCGTGAATATAATAANNNGNNNNACGCTTACATCAACTTGTATGAACAAGAAA 660
QY 740 GATAAAAACATGACCATGATGTTTTAAATTATGTGGTTCGAGCTGATGGCAAAAGAGGC 799
Db 661 GATAAGAAAATAGCTATGACATTTTGAATTATGTGCGGAGAGCTGATGTAAGAAAGAGGC 720
QY 800 TCATGGAGGACTGACAAACAGAGCCAGGAGTGAAGATGAGATGTAACAAAATCTCTGTG 859
Db 721 TCCTGGAGGACTGGTAAACCTGAAGCCAGGAATGAAGATGAATGTATAAAATTTCTCTTG 780
QY 860 AATGATTATGATACCGCAGAGCAGATCCTGTATGGAGAACCGCGAGCTGAAGAGGTG 919
Db 781 AATGATTATGAATATCGTCAGAAACAATCCTAATGGAATATGCAAGAACTTAAAGAGGT 840
QY 920 CTCACGAGATGAAGAAAGAGATGATCTCTCTCTGCTTCCTTCAGAAAGAAAGCCACGG 979
Db 841 CTTCAACAAATGAAAAAGAAATGATTTCTCTCTCTTCTCCCAAAAGAAAGAACCTAGA 900
QY 980 GAAGAGCAGAGGACGGCAGACGCTGCTGCTATCTCGATATAGAAGATGACTCTGG 1039
Db 901 GAAGAGTAGATGATAGTACAGAACTGT---TATTTCTGATGTTGAAGAGATGCCGG 957
QY 1040 GAACTGAGCAGACAGCGGTGTGGGCGCTTTCTCTGTGACACTGTGAGAGAGCAGCTGACA 1099
Db 958 GAACCTAAGCAGAGAGATGATGTGGACCTTCTCTGTGAACCTGTGAGAGAGCAGCTTACA 1017
QY 1100 AACAGCATCAGAAACAGTGGAGAAATTTTGAAGTCTATGTAGAAAACTCGATAACCAA 1159
Db 1018 AACAGCATCAGAAACAGTGGAGAAATTTTGAAGTCTATGTAGAAAACTTGTATAACCAA 1077
QY 1160 GCTTCGAGGTACTACAGAGGCTTAAAGAGGAGCGTCATCTCAGCAGACAGACCAT 1219
Db 1078 GTTTCAAAGGTACACCTGGAAGGTTTAAATGATGAAGATGTAATCTCAGCAGACAGCAT 1137
QY 1220 GAGCAAGAGACTGAGAAACTGAGCTGGAGATTTGAGCGGTGTAAAGAGATGATCAAGCT 1279
Db 1138 GACCAAGAACTGAAAACTCGATTAGAAATTCAGAGGTGTAAAGAAATGATTAAGCT 1197
QY 1280 CAGCAGCAGCTCTTACAGCAGAGCTGCCACCGCTGTGATGATGACACACCTCACTG 1339
Db 1198 CAGCAACAACCTTTACAGCAGCAGCTCGCTACTGCTGATGATGATGATACCACTCACTA 1257
QY 1340 TTGCGAGACTGTACTTCTGGAAGAAAGAAAGCCCTTAAAGAGAGAGTGACCCCTTTT 1399
Db 1258 TTACGAGACTGTATTGTTGGAAGAAAGAAAGCGCTCTCAAGAGAGAAATGGTCCCTTTT 1317
QY 1400 AAAGAGCAAAAAAGAAATTTTGAAGAGAAAGGCGGAGCTTTTACAGAGCTGCCATTCGA 1459
Db 1318 AAAGAGCAAAAAAGAAATTTTGAAGAGGAGAGACGAGCTTTTACAGAGCGCTATTTCG 1377
```

```
QY 1460 TTGGGCTTGGAGAGAAAGCGCTTTGAAGAAGACGAGCCAGCTGGTAAAGCAGCAGTTT 1519
Db 1378 CTGGGATTTGGAGAGAAAGCGCTTTGAAGAAGAAAGCAGCTGGTAAAGCAGCAGTTT 1437
QY 1520 TTAAATCATGACGAACCTTTGACCCAGACAACTCAGAAAATGTGAAACTTTTCAGTGCCTTC 1579
Db 1438 TTAAATATGACTACTACCTTTGACCCAGACAACTCAGAAAATGTGAAACTTTTCAGTGCCTTC 1497
QY 1580 TCAGGAAGTCTTGATCCAGACAACTTATAGTCCACTCACGCCACGGCAAAAGAGCTA 1639
Db 1498 TCAGGAAGTCTTGATTTGGACAACTTATAGTGCACTCGAGCAGCCGCAAAAGAGCCT 1557
QY 1640 CACAGTGTGGCTTAATGGGGTCCAGCTTGCACATCAAAACTGACTAAATCTCTTCTGCTC 1699
Db 1558 CACAGTGTGTCTAATGGGTCTCCAGTTTGCATGTCTAAACTTACTAAATCTCTTCTGCTC 1617
QY 1700 TCACCTTCTACTTCAGACTTTTCCGACAGACAACTTATGCTGTCTGAAACATATGTTCAATC 1759
Db 1618 TCACCTTCCACTTCAGACTTTTGGCCAGACACGTTCTGTCATGTCTGAAACATATGTTCAATC 1677
QY 1760 AGTGTGCTGAATATAAATCCTGAAAGAAAGTAAACCAAGTGAAGTTGCAAGAGAAAGCAG 1819
Db 1678 AATGTACTGAATATAAATCTGCTGAAGAAATTTAAACCAATCAGTTTGGAGGAGAAATGTACA 1737
QY 1820 GATCAGAAAGTGGAGCGTGCAGTCGAGGCGCCAGCTCGCGGAGGGGTGCTACAGCGGATGC 1879
Db 1738 AATCAAAATGGAGTGTGGCGTCAAGCTCGATCACAGGAAGGTTGCTATAGTGTGATGC 1797
QY 1880 TCCTCGGCTTCAGGAGCGCTCAGCGGACCGAGATGACTTACCTTA 1926
Db 1798 TCCTTGACTACAAATTTCTCATGTAGAAAAAGATGACTTACCTTA 1844
```

RESULT 11

```
CX565842
LOCUS CX565842
DEFINITION UI-M-IBO-cum-g-09-0-UI.r1 NIH_BMAP_IBO Mus musculus cDNA clone
IMAGE:30945752 5', mRNA sequence.
ACCESSION CX565842
VERSION CX565842.1 GI:57592871
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 878)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Straubeberg, Ph.D.
Email: csapbe-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
```

FEATURES

```
Seq primer: pYX-5.
Location/Qualifiers
1..878
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30945752"
/tissue_type="whole eye"
/dev_stage="newborn (1,5,15 days)"
/lab_hosts="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_IBO"
```

/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AATTAATACG. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)'. "

ORIGIN		Query Match	32.6%	Score 876.4	DB 9	Length 878	
		Best Local Similarity	99.9%	Pred. No. 7.9e-215			
		Mismatches	877	Conservative	0	Mismatches	1
		Indels	0	Gaps	0		
Qy	388	GAACGAGCTGCTGCTTACGCGAAGAACTGCTGCCAGGAGAGCGTGGAGACACA	447				
Db	1	GAACGAGCTGCTGCTTACGCGAAGAACTGCTGCCAGGAGAGCGTGGAGACACA	60				
Qy	448	GAACCTTGAAGCTGGGACGTGACATGGACCACTGCAGAGCTGCTACGCCAACTTAAGGA	507				
Db	61	GAACCTTGAAGCTGGGACGTGACATGGACCACTGCAGAGCTGCTACGCCAACTTAAGGA	120				
Qy	508	GCAGTTGAAACGCTCCAGCGGGAGATGATCGGGCTTCAAGAGAGACAGCGAGCTGCA	567				
Db	121	GCAGTTGAAACGCTCCAGCGGGAGATGATCGGGCTTCAAGAGAGACAGCGAGCTGCA	180				
Qy	568	GTGCAGAACAGGAGTTTGATCAGCTCTCTGAAGATGAGAAAGATGAGGTACAAAATT	627				
Db	181	GTGCAGAACAGGAGTTTGATCAGCTCTCTGAAGATGAGAAAGATGAGGTACAAAATT	240				
Qy	628	ACAAATATCATAGCCAGCGGGCTACTCAGTAAATCATGATGTAAGAGGAAGGAGCG	687				
Db	241	ACAAATATCATAGCCAGCGGGCTACTCAGTAAATCATGATGTAAGAGGAAGGAGCG	300				
Qy	698	TGAATATATTAAGCTAAAGAGCGCGCTGCATCAGCTCGTTATGAACAAGAGGATAAAA	747				
Db	301	TGAATATATTAAGCTAAAGAGCGCGCTGCATCAGCTCGTTATGAACAAGAGGATAAAA	360				
Qy	748	CATAGCCATGATGTTTTAAATATTATGTTGGTTCGAGCTGATGCCAAACGAGCTCATGGAG	807				
Db	361	CATAGCCATGATGTTTTAAATATTATGTTGGTTCGAGCTGATGCCAAACGAGCTCATGGAG	420				
Qy	808	GACTGACAAAACAGAGCCAGGATGAAGATGAGATGTACAAAATTCGTTGAATGATTA	867				
Db	421	GACTGACAAAACAGAGCCAGGATGAAGATGAGATGTACAAAATTCGTTGAATGATTA	480				
Qy	868	TCAGTACCGCCAGAGCAGATCTCTGATGGAGAACCGGAGCTGAAGAGGTCTCCAGCA	927				
Db	481	TCAGTACCGCCAGAGCAGATCTCTGATGGAGAACCGGAGCTGAAGAGGTCTCCAGCA	540				
Qy	928	GATGAAGAGAGAGATGATCTCTCTCTCTCCTCAGAAAGAGAGCCCGGAAAGAGC	987				
Db	541	GATGAAGAGAGAGATGATCTCTCTCTCTCCTCAGAAAGAGAGCCCGGAAAGAGC	600				
Qy	988	AGAGGACGGCACAGGCACTGTTGCTATCTCCGATATAGAAGTACTCTGGGGAACTGAG	1047				
Db	601	AGAGGACGGCACAGGCACTGTTGCTATCTCCGATATAGAAGTACTCTGGGGAACTGAG	660				
Qy	1048	CAGAGACAGCGTGTGGGCCCTTTCTGTGACACATGTGTGAGAGAGCAGCTGACAAACAGCAT	1107				
Db	661	CAGAGACAGCGTGTGGGCCCTTTCTGTGACACATGTGTGAGAGAGCAGCTGACAAACAGCAT	720				
Qy	1108	CAGGAAACAGTGGAGAAATTTTGAAGATCATGTAGAAAATTCGATACCAAGCTTCGAA	1167				
Db	721	CAGGAAACAGTGGAGAAATTTTGAAGATCATGTAGAAAATTCGATACCAAGCTTCGAA	780				
Qy	1168	GGTACACTCAGAGGGCCCTTAATGAGGAGGAGCTCATCTCAGGACAAAGACCATGAGCAAGA	1227				

Db	781																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
----	-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

```

Db      61  AATGAGGAGGAGCTCATCTCAGCACAAGACCATGACCAAGAGACTGAGAAACTGGAGCTG 120
QY      1247 GAGATTGACGGGTGTTAAAGAGATGATCAAGGCTCAGCAGAGCTCTTACAGAGCAGCTG 1306
Db      121  GAGATTGACGGGTGTTAAAGAGATGATCAAGGCTCAGCAGAGCTCTTAA---CAGCAGCTG 177
QY      1307 GCACCACTGTGATGATGACACCACTCCTCACTGTTGCGAGACTGTTACTTGTGGAAGAA 1366
Db      178  GCCACCACTGTGATGATGACACCACTCCTCACTGTTGCGAGACTGTTACTTGTGGAAGAA 237
QY      1367 AAGGAACGCTTTAAAGAGAGTGGACCTTTTAAAGAGCAAAAGAAATTTGAGAGA 1426
Db      238  AAGGAACGCTTTAAAGAGAGTGGACCTTTTAAAGAGCAAAAGAAATTTGAGAGA 297
QY      1427 GAAAGCGGAGCTTTTACAGAACTGTCATTCGATTGGGTTGGAGAGAAAGCGTTTGA 1486
Db      298  GAAAGCGGAGCTTTTACAGAACTGTCATTCGATTGGGTTGGAGAGAAAGCGTTTGA 357
QY      1487 GAAGAGCGAGCCAGCTGGGTAAGCAGCAGTGTGTTTAAACATGACCAACTTTGACCAACCAG 1546
Db      358  GAAGAGCGAGCCAGCTGGGTAAGCAGCAGTGTGTTTAAACATGACCAACTTTGACCAACCAG 417
QY      1547 AACTCAGAAATGTGAAACTTTTCAAGTGCCTTCTCAGGAAGTTCGATCCAGACAATCTT 1606
Db      418  AACTCAGAAATGTGAAACTTTTCAAGTGCCTTCTCAGGAAGTTCGATCCAGACAATCTT 477
QY      1607 ATAGTCCACTCAGCGCCACGCAAGAAAGCTACACAGTGTGGCTAAATGGGTGCGAGCT 1666
Db      478  ATAGTCCACTCAGCGCCACGCAAGAAAGCTACACAGTGTGGCTAAATGGGTGCGAGCT 537
QY      1667 TGCACATCAAAACTGACTAAATCTCTCTGCTGCTCAGTGTGCTGAATATAACTCCTGAAGAA 1786
Db      538  TGCACATCAAAACTGACTAAATCTCTCTGCTGCTCAGTGTGCTGAATATAACTCCTGAAGAA 1786
QY      1727 ACACATTCATGTGTCTGTAACACACAGTTCCTCAGTGTGCTGAATATAACTCCTGAAGAA 1786
Db      598  ACACATTCATGTGTCTGTAACACACAGTTCCTCAGTGTGCTGAATATAACTCCTGAAGAA 657
QY      1787 AGTAAACCAAGTGAAGTTCGAAGAAAGCAGGATCAAGTGAAGTGAAGTGAAGTGAAGT 1846
Db      658  AGTAAACCAAGTGAAGTTCGAAGAAAGCAGGATCAAGTGAAGTGAAGTGAAGTGAAGT 717
QY      1847 CCCAGCTCGCGGAGGGTGTACAGCGGATGCTCCTCGGCTTCAGGAGCGCTCACGG 1906
Db      718  CCCAGCTCGCGGAGGGTGTACAGCGGATGCTCCTCGGCTTCAGGAGCGCTCACGG 777
QY      1907 GACCGAGATGACTTACCTTAAATGTGCGGGTGCAGTGTGTTCCAGATGTGCGCTAGA 1966
Db      778  GACCGAGATGACTTACCTTAAATGTGCGGGTGCAGTGTGTTCCAGATGTGCGCTAGA 836
QY      1967 GGAG 1970
Db      837 GGAG 840

```

```

RESULT 13
CX562093
LOCUS      CX562093
DEFINITION  CX562093 834 bp mRNA linear EST 12-JAN-2005
IMAGE:30699705 5', mRNA sequence.
ACCESSION  CX562093
VERSION    CX562093.1 GI:57589122
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 834)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.

```

Email: cgapbe-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mouseefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

FEATURES

Location/Qualifiers
1..834
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30699705"
/tissue_type="whole brain"
/dev_stages="embryo 12.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_FCO"
/note="Organ: brain; Vector: pYX-Asc; Site 1: Ecor I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dr primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction.
ligated with Ecor I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA
tail, is TGAGAGAGCC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP). 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 30.1%; Score 810; DB 9; Length 834;
Best Local Similarity 99.5%; Pred. No. 1.2e-197;
Matches 832; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

```

QY      346  GCGAAGAGAGAAATTAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTCGTGT 405
Db      1    GCGAAGAGAGAA-TAAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTCGTGT 59
QY      406  TCAGCGGAAGAACCTGCTGGCCAGGAGAGCTGGAGACACAGAACTTGAAGCTGGGAG 465
Db      60  TCAGCGGAAGAACCTGCTGGCCAGGAGAGCTGGAGACACAGAACTTGAAGCTGGGAG 119
QY      466  TGACATGGACCACTGCAGAGCTGCTACGCCAAACTTAAGGAGCAGTTCGAAACCTCCAG 525
Db      120  TGACATGGACCACTGCAGAGCTGCTACGCCAAACTTAAGGAGCAGTTCGAAACCTCCAG 179
QY      526  GCGGAGATGATCGGGCTTCAAGAGAGAGACAGGAGCTGCGAGTGCACAGACAGAGTTT 585
Db      180  GCGGAGATGATCGGGCTTCAAGAGAGAGACAGGAGCTGCGAGTGCACAGACAGAGTTT 239
QY      586  GCATCAGCTCTGAAGATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAG 645
Db      240  GCATCAGCTCTCTGAAGATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAG 299
QY      646  CCGGGCTACTCAGTATATCATGTGTGAAGAGGAGGAGCGGTGAATATAAAGCTAAA 705
Db      300  CCGGGCTACTCAGTATATCATGTGTGAAGAGGAGGAGCGGTGAATATAAAGCTAAA 359
QY      706  GGAGCGCTTCAGCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTT 765
Db      360  GGAGCGCTTCAGCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTT 419
QY      766  AAATTATGTGGTTCAGCTGATGCAACAGGCTCATGGAGGACTGCAAAAACAGAGCC 825
Db      420  AAATTATGTGGTTCAGCTGATGCAACAGGCTCATGGAGGACTGCAAAAACAGAGCC 479

```

826 CAGGAATGAAGATGAGATGTACAAATCTCTTGAATGATTATGATACCGCCAGAGCA 885
 |||||
 480 CAGGAATGAAGATGAGATGTACAAATCTCTTGAATGATTATGATACCGCCAGAGCA 539
 |||||
 886 GATCCTGATGAGAAACGGGAGCTGAAGAGGTCCTCCAGCAGATGAAGAGGAGATGAT 945
 |||||
 540 GATCCTGATGAGAAACGGGAGCTGAAGAGGTCCTCCAGCAGATGAAGAGGAGATGAT 599
 |||||
 946 CTCTCTCTGTCCTCTCAGAAAGAAAGCCAGGGAAGAGCAGAGGACGCGACAGGCAC 1005
 |||||
 600 CTCTCTCTGTCCTCTCAGAAAGAAACCCAGGGAAGAGCAGAGGACGCGACAGGCAC 658
 |||||
 1006 TGTTCCTATCTCCATATAGAGATGACTCTGGGGAAGTGAAGCAGACAGCGTGTGGGG 1065
 |||||
 659 TGTTCCTATCTCCATATAGAGATGACTCTGGGGAAGTGAAGCAGACAGCGTGTGGGG 718
 |||||
 1066 CTTTTCCTGTGACACTGTGAGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAT 1125
 |||||
 719 CTTTTCCTGTGACACTGTGAGAGAGCAGCTGACANACAGCATCAGGAAACAGTGGAGAT 778
 |||||
 1126 TTTGAAAGTCATGTAGAAAACTCGATAACCAAGCTTCGAAGTACACTCAGAGG 1181
 |||||
 779 TTTGANAAGTCATGTAGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGG 834
 |||||

RESULT 14

CN539511

LOCUS

DEFINITION UI-M-HUO-cqu-n-01-0-UI.r1 NIH_BMAP_HUO Mus musculus cDNA clone
 IMAGE:30667224 5', mRNA sequence.

CN539511

VERSION

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 804)

NIH-MGC

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. James Lin University of Iowa

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/mousefl.html

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1..804

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:30667224"

/tissue_type="whole eye"

/dev_stages="newborn(1, 5, 15 days)"

/lab_hosts="DH10B (T1 phage resistant)"

/clone_lib="NIH_BMAP_HUO"

/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;

Site 2: Not I; The library was constructed according

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. Denatured RNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with oligo-dT

primer containing a Not I site. Double strand cDNA was

size selected according to mRNA size fraction, ligated

with EcoR I adaptor, digested with NotI and then cloned

directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AATAATACG. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH)."

ORIGIN

Query Match 29.8%; Score 802; DB 8; Length 804;
 Best Local Similarity 99.8%; Pred. No. 1.3e-195;
 Matches 802; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 646 CCGGGCTACTCAGTATATCATGTGAAGAGAGGAGCGTGAATATATAAGCTAAA 705
 |||||
 Db 1 CCGGGCTACTCAGTATATCATGTGAAGAGAGGAGCGTGAATATATAAGCTAAA 60
 |||||
 Qy 706 GGAGCGCTCATCAGCTCGTTATGAACAAGAGAGTAAACATACGCCATGGATGTTT 765
 |||||
 Db 61 GGAGCGCTCATCAGCTCGTTATGAACAAGAGAGTAAACATACGCCATGGATGTTT 120
 |||||
 Qy 766 AAATTATGTTGGTTCGAGCTGATGCGAAACGAGGCTCATGAGGACTGACAAACAGAGC 825
 |||||
 Db 121 AAATTATGTTGGTTCGAGCTGATGCGAAACGAGGCTCATGAGGACTGACAAACAGAGC 180
 |||||
 Qy 826 CAGGAATGAAGATGAGATGTACAAAATTCTGTTGAATGATTATGAGTACCGCCAGAGCA 885
 |||||
 Db 181 CAGGAATGAAGATGAGATGTACAAAATTCTGTTGAATGATTATGAGTACCGCCAGAGCA 240
 |||||
 Qy 886 GATCCTGATGAGAACCGGAGCTGGAAGAGGTCCTCCAGCAGATGAAGAGGAGATGAT 945
 |||||
 Db 241 GATCCTGATGAGAACCGGAGCTGGAAGAGGTCCTCCAGCAGATGAAGAGGAGATGAT 300
 |||||
 Qy 946 CTCTCTCTCTCTCTCAGAGAGAGCCAGGGAAGAGCAGAGGACGCGCAGAGCAC 1005
 |||||
 Db 301 CTCTCTCTCTCTCTCAGAGAGAGCCAGGGAAGAGCAGAGGACGCGCAGAGCAC 360
 |||||
 Qy 1006 TGTTCCTATCTCCGATATAGAAGATGACTCTCGGGAACTCAGCAGAGACAGCGTGTGGGG 1065
 |||||
 Db 361 TGTTCCTATCTCCGATATAGAAGATGACTCTCGGGAACTCAGCAGAGACAGCGTGTGGGG 420
 |||||
 Qy 1066 CTTTTCCTGTGACACTGTGAGAGAGCAGCTGACAAACAGCATCAGGAAACAGTGGAGAT 1125
 |||||
 Db 421 CTTTTCCTGTGACACTGTGAGAGAGCAGCTGACAAACAGCATCAGGAAACAGTGGAGAT 480
 |||||
 Qy 1126 TTTGAAAGTCATGTAGAAAACTCGATACCAAGCTTCGAAGGTACACTCAGAGGCGCT 1185
 |||||
 Db 481 TTTGAAAGTCATGTAGAAAACTCGATACCAAGCTTCGAAGGTACACTCAGAGGCGCT 540
 |||||
 Qy 1186 TAATGAGGAGGAGCTCATCTCCAGCAAGACCATGAGCAAGAGACTGAGAAACTGGAGCT 1245
 |||||
 Db 541 TAATGAGGAGGAGCTCATCTCAGCAAGACCATGAGCAAGAGACTGAGAAACTGGAGCT 600
 |||||
 Qy 1246 GGAGATTGAGCGGTGTTAAAGAGATGATCAAGGCTCAGCAGAGCTCTTACAGCAGCAGCT 1305
 |||||
 Db 601 GGAGATTGAGCGGTGTTAAAGAGATGATCAAGGCTCAGCAGAGCTCTTACAGCAGCAGCT 660
 |||||
 Qy 1306 GGCCACACAGCTGTGATGATGACACACCCTCAGCTTTCGGAGACTGTTACTTCTGGAGGA 1365
 |||||
 Db 661 GGCCACACAGCTGTGATGATGACACACCCTCAGCTTTCGGAGACTGTTACTTCTGGAGGA 720
 |||||
 Qy 1366 AAAGGAACCGCTTAAAGAGAGTGGACCCCTTTTAAAGAGCAAAAAAGAAATTTTGAGAG 1425
 |||||
 Db 721 AAAGGAACCGCTTAAAGAGAGTGGACCCCTTTTAAAGAGCAAAAAAGAAATTTTGAGAG 780
 |||||
 Qy 1426 AGAAAGCGGAAGCTTTTACAGAGC 1449
 |||||
 Db 781 AGAAAGCGGAAGCTTTTACAGAGC 804
 |||||

RESULT 15

CX562859

LOCUS

DEFINITION UI-M-IB0-cub-p-21-0-UI.r1 NIH_BMAP_IB0 Mus musculus cDNA clone

798 bp

mrna

linear

EST 12-JAN-2005

IMAGE:6807142 5', mRNA sequence.
 ACCESSION CX562859
 VERSION CX562859.1 GI:57589888
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridea; Muridae; Murinae; Mus.
 1 (bases 1 to 798)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. James Lin University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousef1.html>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

Seq primer: pYX-5.
 Location/Qualifiers
 1..798
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="CS7BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6807142"
 /tissue_type="whole eye"
 /dev_stages="newborn (1,5,15 days)"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP IB0"
 /note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AATAATACG. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH)."

ORIGIN

Query Match 29.6%; Score 798; DB 9; Length 798;
 Best Local Similarity 100.0%; Pred. No. 1.4e-194;
 Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 150 CAGAAACAAGATGTCTCCGTCAGTTGTACTCCAGCAAGTTCTGTGCTCTTCAGTAC 209
 DB 1 CAGAAACAAGATGTCTCCGTCAGTTGTACTCCAGCAAGTTCTGTGCTCTTCAGTAC 60
 QY 210 CTTTATCCAAAACGTCATGCTGTTTTTCGGTGTCTTCTGCACAGGAGAGAACATTGAAC 269
 DB 61 CTTTATCCAAAACGTCATGCTGTTTTTCGGTGTCTTCTGCACAGGAGAGAACATTGAAC 120
 QY 270 AAAGTATTTCTTATCTTTGATCAGGAGCTGACACCTTCGGGTTTTCTTCTTGTATGAAG 329
 DB 121 AAAGTATTTCTTATCTTTGATCAGGAGCTGACACCTTCGGGTTTTCTTCTTGTATGAAG 180
 QY 330 AATCCAAAAGTAAAGAGCAAGAGAGAAATTAATATAGTCGCTGTTCTGAATGTATGA 389
 DB 181 AATCCAAAAGTAAAGAGCAAGAGAGAAATTAATATAGTCGCTGTTCTGAATGTATGA 240
 QY 390 ACGAGCTGCTGCTCTTTCAGCGGAAGAACCTGCTGCGCCAGGAGAGCGTGGAGACACAGA 449

DB 241 AGGAGCTGCTGCTGCTTTCAGCGGAAGAACCTGCTGCGCCAGGAGAGCGTGGAGACACAGA 300
 QY 450 ACTTGAAGCTGGGCGAGTGACATGGACCACTGTCAGAGCTGTCTACGCCAAACTTTAAGGAGC 509
 DB 301 ACTTGAAGCTGGGCGAGTGACATGGACCACTGTCAGAGCTGTCTACGCCAAACTTTAAGGAGC 360
 QY 510 AGTTGGAACCTCCAGGCGGAGAGATGATCGGGCTTCAAGAGAGAGACAGGCGAGCTGCAGT 569
 DB 361 AGTTGGAACCTCCAGGCGGAGAGATGATCGGGCTTCAAGAGAGAGACAGGCGAGCTGCAGT 420
 QY 570 GCAAGAACAGGAGTTTGGTCATCAGCTCCCTGGAAGAATGAGAAAGATGAGGTACAAAAATTAC 629
 DB 421 GCAAGAACAGGAGTTTGGTCATCAGCTCCCTGGAAGAATGAGAAAGATGAGGTACAAAAATTAC 480
 QY 630 AAAATATCATAGCCAGCGGGCTACTCAGTATATAATCATGATGTGAAGAGGAGGAGCGTG 689
 DB 481 AAAATATCATAGCCAGCGGGCTACTCAGTATATAATCATGATGTGAAGAGGAGGAGCGTG 540
 QY 690 AATATAAAGCTAAAGGAGCGCTGCATCAGCTCGTTATGAAACAGAGGATAAAAACA 749
 DB 541 AATATAAAGCTAAAGGAGCGCTGCATCAGCTCGTTATGAAACAGAGGATAAAAACA 600
 QY 750 TAGCCATGGATGTTTTTAAATTATGTGGCTCGAGCTGTGGCAACGAGGCTCATCGAGGA 809
 DB 601 TAGCCATGGATGTTTTTAAATTATGTGGCTCGAGCTGTGGCAACGAGGCTCATCGAGGA 660
 QY 810 CTGACAAAACAAGAGCCAGGAATGAAGATGAGATGTACAAAATTCTGTTGAATGATTATG 869
 DB 661 CTGACAAAACAAGAGCCAGGAATGAAGATGAGATGTACAAAATTCTGTTGAATGATTATG 720
 QY 870 AGTACCCGCGAAGCAGATCCTGATGGAGAACCGCGGAGCTGGAAGAGTCTCTCCAGCAGA 929
 DB 721 AGTACCCGCGAAGCAGATCCTGATGGAGAACCGCGGAGCTGGAAGAGTCTCTCCAGCAGA 780
 QY 930 TGAAGAAAGGAGATCATCT 947
 DB 781 TGAAGAAAGGAGATCATCT 798

Search completed: June 14, 2006, 04:03:37
 Job time : 11424 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 13, 2006, 23:24:10 ; Search time 7913 Seconds
(without alignments)
6519.093 Million cell updates/sec

Title: US-10-644-084-2

Perfect score: 3165

Sequence: 1 MGDWMTVPVLTENKNLS.....CYSGCSSAFSAHGDRDDLP 615

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh
-Q/abses/ABSSWEB spool/US10644084/runat_12062006_150418_1399/app query.fasta_1
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LISP=45
-DOCALIGN=200 -THR SCORE=act -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abses08
-USER=US10644084 @CGN_1_1_6323 @runat_12062006_150418_1399 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est3:*
3: gb_est4:*
4: gb_est5:*
5: gb_est6:*
6: gb_hcc:*
7: gb_est2:*
8: gb_est7:*
9: gb_est8:*
10: gb_est9:*
11: gb_gse1:*
12: gb_gse2:*
13: gb_gse3:*
14: gb_gse4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3150.5	99.5	1845	14 AY412493	AY412493 Mus muscu
2	3149.5	99.5	3185	6 AK043865	AK043865 Mus muscu
3	3149.5	99.5	3413	6 AK031356	AK031356 Mus muscu
4	3140.5	99.2	2245	6 AK077112	AK077112 Mus muscu

5	3138	99.1	2371	6	AK049080
6	2786.5	88.0	1845	14	AY412491
7	2786.5	88.0	2111	6	CR613292
8	2786.5	88.0	3154	6	BC035580
9	2786.5	88.0	5658	6	BSM801317
10	2747.5	86.8	1845	14	AY412492
11	1482	46.8	878	9	CX565842
12	1416.5	44.8	903	10	DV893190
13	1364	43.1	804	8	CNS39511
14	1356	42.8	834	9	CX562093
15	1349.5	42.6	840	5	CF949831
16	1348	42.6	798	9	CX562859
17	1323.5	41.8	933	4	CB205832
18	1314	41.5	822	8	CN563162
19	1300.5	41.1	891	9	DV286111
20	1291	40.8	771	3	BQ179212
21	1287	40.7	860	10	DT831879
22	1263	39.9	770	5	CF729394
23	1251	39.5	760	8	CNS28866
24	1235	39.0	746	8	CNS28831
25	1233.5	39.0	845	3	BQ769157
26	1221	38.6	723	5	CF737957
27	1216.5	38.4	1005	1	AL582610
28	1216	38.4	828	10	DV803752
29	1208	38.2	712	5	CF732776
30	1199.5	37.9	800	9	DN877534
31	1196.5	37.8	833	8	CV117234
32	1192	37.7	825	1	AU133150
33	1184	37.4	954	1	AL561392
34	1168	36.9	736	8	CX226608
35	1162	36.7	930	4	BX372332
36	1153	36.4	771	5	CF737129
37	1149	36.3	687	9	CX568194
38	1145	36.2	679	8	CNS32570
39	1141	36.1	683	8	CO424815
40	1140.5	36.0	733	9	DN864383
41	1076.5	34.0	805	2	BG242849
42	1075.5	34.0	638	8	CN720771
43	1069	33.8	751	5	CJ465026
44	1062	33.6	640	5	CF737822
45	1056	33.4	748	8	CO555994

ALIGNMENTS

RESULT 1	AY412493	1845 bp	DNA	linear	GSS 12-DEC-2003
LOCUS	AY412493	Mus musculus HCM4575 gene, VIRTUAL TRANSCRIPT, partial sequence,			
DEFINITION	AY412493	genomic survey sequence.			
ACCESSION	AY412493	GI:39768458			
VERSION	AY412493.1	GI:39768458			
KEYWORDS	GSS.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.				
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
PUBMED	14671302				
REFERENCE	2 (bases 1 to 1845)				
AUTHORS	Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.				
TITLE	Direct Submission				

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..1845
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
gene <!.>1845
/locus_tag="HCM4575"
ORIGIN
Alignment Scores:
Pred. No.: 3,266-304 Length: 1845
Score: 3150.50 Matches: 614
Percent Similarity: 99.8% Conservative: 0
Best Local Similarity: 99.8% Mismatches: 0
Query Match: 99.5% Indels: 1
DB: 14 Gaps: 1
US-10-644-084-2 (1-615) x AY412493 (1-1845)
QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
DB 1 ATGGGAGATTGGATGACTGTGACAGATCCAGTCTCTGTGTACAGAAAACAAAAATCTCT 60
QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnValLeuCys 40
DB 61 CAATATACCTCGAAGAACAGATGTCCTCGTCCAGTTTGTACTCCAGCAAGTCTGTGC 120
QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
DB 121 TCTTCAGTACCTTTATCCAAAACGTCATGTGTGTTTCGGGTCTTCTGCACAGAGAG 180
QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
DB 181 AACATTGAACAAAGTATTCTCTATCTTGCATCAGGAGCTGACCACCTTCGGGTTCCTCC 240
QY 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
DB 241 TTGTATGAAGAAATCCAAAAGTAAAGAGGCAAGAGAGAAATTAATATAGTCGCTGTTCTG 300
QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
DB 301 AACTGTATGAACGACTGCTGCTGCTTTCAGCGGAAGAACCTGCTGCTCCAGAGAGCGGTG 360
QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
DB 361 GAGACACAGAACTTGAAGCTGGGCGAGTGACATGGACCACCTGCAGAGCTGCTACGCCAAA 420
QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
DB 421 CTTAAGGAGCAGTTGGAAACGTCACGCGGGAGATGATCGGGCTTCAAGAGAGACACAGG 480
QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
DB 481 CAGCTGCAGTGCAGAACAGGAGTTTGCATCAGCTCTCTGAAGAAATGAGAAAGATGAGGTA 540
QY 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
DB 541 CAAAAATACAAAATATCATGCCAGCGGGCTACTCAGTATAATATGATGTGAAGAGG 600
QY 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
DB 601 AAGGAGCGGTGAATATAAATAAGCTAAAGGAGCGCCCTGCATCAGCTCGTTATGAACAAGAG 660
QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
DB 661 GATAAANAACATGACCATGATGTTTAAATATGTTGGTCTGAGCTGATGGCNAACAGGC 720
QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260

DB 721 TCATGGAGGAGCTGACAAAACAGAACCCAGGAATGAAGATGAGATGTACAAAAATTCGTGTG 780
QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
DB 781 AATGATTATGATACCGCCAGAACAGATCTCTGATGGAGAACCGGAGCTGAAGAGGTC 840
QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
DB 841 CTCAGCAGATGAGAGAGAGATGATCTCTCTCTCTCTCTCTCAGAGAGAGAGGCCAGG 900
QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
DB 901 GAAAGAGCAGAGGACGCGCACAGGCACCT--GCTATCTCCGATATAGAAGATGACTCTGGG 957
QY 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
DB 958 GAACGTGACAGACAGACGCGTGTGGGCGCTTTCCTGTGTGACACTGTGTGAGAGAGCAGCTGACA 1017
QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
DB 1018 AACAGCATCAGAAAACAGTGGAGAAATTTTGAAGAGTCATGTAGAAAAAATCGATAACCAA 1077
QY 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
DB 1078 GCTTCGAAGGTACACTCAGAGGCGCTTAATGAGGAGGACGTCATCTCACACACAGACCAT 1137
QY 381 GluGlnGluThrGluLysLeuGluLeuGluLeuGluArgCysLysGluMetIleLysAla 400
DB 1138 GAGCAAGAGACTGAGAACTGGAGCTGGAGATTGAGCGGTGTAAAGAGATGATCAAGGCT 1197
QY 401 GlnGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
DB 1198 CAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGCTGTGTATGATGATGATGATGATGATG 1257
QY 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440
DB 1258 TTGGCAGACTGTTACTTGTCTGGAAGAAAGAAAGCGCTTAAAGAGAGAGTGGACCCCTTTT 1317
QY 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
DB 1318 AAAGAGCAAAAAAGAAATTTTGAGAGAGAAAGGCGAGCTTTACAGAAAGCTGCCAATCGA 1377
QY 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnPhe 480
DB 1378 TTGGGTGGAGAGAAAGCGCTTTGAAGAGAGCGAGCGACGCTGGGTAAAGCAGCAGTTT 1437
QY 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
DB 1438 TTAACATGACGAACTTTTGACCCACAGAACTCAGAAAAATGTGAAACTTTTTCAGTGCCTTC 1497
QY 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
DB 1498 TCAGAAAGTCTGATCCAGACAACTTATATGTCACCTCAGCGCCACGCGCAAAAGAGCTA 1557
QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
DB 1558 CACAGTGTGGCTAATGGGTGCCAGCTTGACATCAAAACTGACTAAATCTCTCTCTGCC 1617
QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
DB 1618 TCACCTTCTACTCTCAGACTTTTCGCCAGACACATTCATGTGTCTGTAACACACATTTCCATC 1677
QY 561 SerValLeuAsnIleThrProGluLeuSerLysProSerGluValAlaArgGluSerThr 580
DB 1678 AGTGTGCTGAATATAACTCTCTGAAGAAAGTAAACCAAGTGAAGTTGCAAGAGAGAACGACG 1737
QY 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCys 600
DB 1738 GATCAGAAAGTGGAGCGTGCAGTCCAGGCGCCAGCTCGCGGGAGGGGTGCTACAGCGGATGC 1797
QY 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
DB 1798 TCCTCGGCTTTTCAGGAGCGCTCACGGGACCGAGATGACTTACCT 1842

RESULT 2
AK043865
LOCUS
DEFINITION
AK043865 3185 bp mRNA linear HTC 02-SBP-2005
Mus musculus 10 days neonate cortex cDNA, RIKEN full-length
protein library, clone:A830043F14 product:HYPOHETICAL 71.0 KDA
PROTEIN homolog [Mus musculus], full insert sequence.

ACCESSION
AK043865
VERSION
AK043865.1
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
10349636

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
11042159

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tachiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
The FANTOM Consortium, the RIKEN Genome Exploration Research Group
Phase I and II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
RIKEN Genome Exploration Research Group, Genome Science Group
(Genome Network Core Team) and the FANTOM Consortium.
Antisense Transcription in the Mammalian Transcriptome
Science 309, 1564-1566 (2005)

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
The FANTOM Consortium, Riken Genome Exploration Research Group and
Genome Science Group (Genome Network Project Core Group).
The Transcriptional Landscape of the Mammalian Genome
Science 309, 1559-1563 (2005)

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

TITLE
JOURNAL
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa, 230-0045, Japan [E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, tel: 81-45-503-9222,
Fax: 81-45-503-9216]
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN,
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.

FEATURES
Location/Qualifiers
1..3185
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A830043F14"
/db_xref="taxon:10090"
/clone="A830043F14"
/issue_type="cortex"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
389..2233
/note="unnamed protein product; HYPOHETICAL 71.0 KDA
PROTEIN homolog [Mus musculus] (SPTR|AAH21749, evidence:
FASTV, 99.8%ID, 100%length, match=1842)
putative"
/codon_start=1
/protein_id="BAC31684.1"
/db_xref="GI:26335972"
/translation="MGDMVTVDPLCTENKLSQYTSFETKMSPLSYSQOVLCSVP
LSKNVHGVFGCTGEGIEQISLYDQDLTTGFFSLYEESKSKARSLNIVAVLNC
MNEILLQRLNLAQESVETQNLKGDMDHLSQCYAKLKEQLETSSRMIGLQERDR
QLOCKNRSLLHQLKNDEQVQLNIIASRATQYNNHVKRERYNKLRHLQVYNN
KKXNDAMDVLNVVGRADGKRGWRTDKTEARNEDMYKILLNDYERQKQILMENAE
LKVLQOMKKEWISLLSPQKKPRERAEDGTCTVAISDIEDDSGLSRDSVGLSCDT
VREQLTNSIRQWRLKSHVEKLDNQAKVHSEGLNEEDVISRQHQTEKLELBIE
RCKMTAKQAQQLQATTCDDTTSLLRDCLYLLBEKERKEWTLFQKQKNFERER
RSFTAAIRGLERKAFEEERASWYKQFLNMTNFHQNSENVKLSFASGSDPDNL
IVHSRPRQKLHSHVANGVPACTSKLTSLPASPSDFRQTHSCVSEHSSISVLTTP
BESKPEVARESTDOKWSVQSRPSREGCYGCSAFAFRSAHGDRLDP"
3168..3173
/note="putative"
3185
/note="putative"

ORIGIN
Alignment Scores:
Pred. No.: 9, 72e-304 Length: 3185
Score: 3149.50 Matches: 614
Percent Similarity: 99.8% Conservative: 0
Best Local Similarity: 99.8% Mismatches: 0
Query Match: 99.5% Indels: 1
DB: 6 Gaps: 1
US-10-644-084-2 (1-615) x AK043865 (1-3185)

Qy 1 MetGlyAspTTPMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
Db 389 ATGGGAGATTGGATGACTGTGACAGATCCAGTCTGTGTACAGAAAACAAAATCTCTCT 448
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerLeuTyrSerGlnGlnValLeuCys 40
Db 449 CAATATACCTCAGAAACAAAGATGTCCTCGTCCAGTTTGTACTCCACGACAGTTCTGTGC 508
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 509 TCTTCAGTACCTTTTATCAAAACCGTCATGGTGTGTTTTCGGGTGTTCTTCTGCACAGAGAG 568

Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium, the RIKEN Genome Exploration Research Group
Phase I and II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6
RIKEN Genome Exploration Research Group, Genome Science Group
(Genome Network Core Team) and the FANTOM Consortium.
Antisense Transcription in the Mammalian Transcriptome
Science 309, 1564-1566 (2005)
7
The FANTOM Consortium, Riken Genome Exploration Research Group and
Genome Science Group (Genome Network Project Core Group).
The Transcriptional Landscape of the Mammalian Genome
Science 309, 1559-1563 (2005)
8 (bases 1 to 3413)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayato,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Inotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa, 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fanom.gsc.riken.jp/
Location/Qualifiers
1. .3413
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:603041112"
/db_xref="taxon:10090"
/clone="603041112"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="13 days embryo"
264. .2108
/note="unnamed protein product; hypothetical protein
(evidence: ProCrest)
putative"
/codon_start=1
/protein_id="BAC27363.1"

/db_xref="GI:26327239"
/translation="MGDMWTQSDYDPLVTENKNLSQVTSYTKMSPSLYSQQVLCSSVP
LSKNVHGVGFCTGENTIEQISYSDQELTTFGFSFLYESKSEKARELNTVAVLNC
MNEELVLRKNLLAQESVETQNLKLGSDMHLQSCVAKLKEQLTSRRRMIQLQEDR
OLOCKNRLHOLLKNEKEDEVKLONI IASRATQYNHDKRKEYNKLERLHOLVMN
KKDNIAMDVLNVGRADGKRGSMFTDEARNEDEMYKILLNDYEVROKQILMENAE
LKXVLQMKKEMISLPQKXPRERAEDGTGTVAISDIEDSGLSRDSRWGLSCDT
VREQUTNSIRKQWRILKSHVEKLDNOASKVSEGLNEEDVISRQDHQETKLELSE
RCKEMIKAQQLQLLATDQDDTSLLRDCLLLEKERLKEWTLFKQKQKFRER
RSFTEAARIQLGERKAFBEERASWVKQFLNMTNFDHQSENVKLFSFSGSDPNL
I VHSRPROKHLHSVANGVPACTSKLTSLPASPSDPROTHSCVSEHSSISVLNITP
EESKPSVARESTDQKWSVQSRSSREGCYSCSSAFRSAHGRDRLP"
3394. .3399
/note="putative"
3413
/note="putative"
polyA_signal
polyA_site
ORIGIN
Alignment Scores:
Pred. No.: 1.08e-303 Length: 3413
Score: 3149.50 Matches: 614
Percent Similarity: 99.8% Conservat: 0
Best Local Similarity: 99.8% Mismatches: 0
Query Match: 99.5% Indels: 1
DB: 6 Gaps: 1
US-10-644-084-2 (1-615) x AK031356 (1-3413)
Qy 1 MetGlyAspTyrMetThrValThrAspProValLeuLeuCysThrGluAsnLysAsnLeuSer 20
Db 264 ATGGAGATTGGATGACCTGTGACAGATCCAGTTCTGTGTACAGAAAACAAAAATCTCTCT 323
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db 324 CAATATACCTCAGAAACAAAGATGCTCGTCCAGTTGTACTCCAGCAAGTCTGTGTC 383
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 384 TCTTCAGTACCTTTATCCAAAACGTCATGCTGTTGTTGCGGTCTCTCTGTCACAGAGAG 443
Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
Db 444 AACATTGAACAAAGTATTTCTTATCTTGTATCAGGAGCTGACCACCTTCGGGTTCCTTCC 503
Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db 504 TTGTATGAAGAATCCAAAAGTAAACAGGCAAGAGAGAGAAATTAATATAGTCTGCTGTTCTG 563
Qy 101 AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
Db 564 AACTGTATGAACAGCTGCTCGTGTTCAGCGGAAGAACCTCTGCGCCAGGAGAGCGTG 623
Qy 121 GluThrGlnAsnLeuLysGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
Db 624 GAGACACAGAACTTGAAGCTGGCAGTGCATGGACACCTGCAGAGCTGCTACGCCAAA 683
Qy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
Db 684 CTTAAGGAGCAGTTGGAAACGTCACAGCGGAGAGATGATCGGCTTCAAGAGAGACAGG 743
Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
Db 744 CAGCTGCAGTGCAGAACACAGGAGTTTGATCAGCTCTCTGAAGATGAGAAAGATGAGGTA 803
Qy 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Db 804 CAAAATTTACAAAATATCATGCCCGCGCTACTCAGTATATCATGATGTGAAGAGG 863
Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db 864 AAGGAGCGTGAATATAATAAGCTAAAGGAGCGCCTGCATCAGTCTGTTATGAACAAGAAG 923
Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240

Db	924		GATAAAAACATAGCCATCGATGTTTAAATTTATGTGGGTGAGCTGATGGCAACAGAGGC	983
Qy	241	SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyLysIleLeuLeu	260	
Db	984	TCATGGAGGACTGACAAACAGAGCCAGGAATGAGATGAGATGTACAAAATTTCTGTG	1043	
Qy	261	AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal	280	
Db	1044	AATGATTATGATGACCGCCAGAAGCAGATCTCTGATGGAGACCGGAGCTGAAGAAGTC	1103	
Qy	281	LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg	300	
Db	1104	CTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTGCTCTCTCAGAAAGAAAGCCAGG	1163	
Qy	301	GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly	320	
Db	1164	GAAGAGCAGAGGACGGCAGCAGGCATGTTGCTTATCTCCGATATAGAAGATGACTCTGGG	1223	
Qy	321	GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr	340	
Db	1224	GAACCTGAGCAGAGACAGCGTGTGGGCGCTTTCTGTGCACCTGTGAGAGAGCAGCTGACA	1283	
Qy	341	AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln	360	
Db	1284	AACAGCATCAGAAACAGTGAGAGAAATTTTGAAAAAGTCATGTAGAAAAAATCGATAACCA	1343	
Qy	361	AlaSerLysValHisSerGluGlyLysAsnGluLysAspValIleSerArgGlnAspHis	380	
Db	1344	GCCTCGAAGGTACACTCAGAGGGCGCTTAATGAGAGGACGTCATCTCACGACAAAGACCAT	1403	
Qy	381	GluGlnGlnThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla	400	
Db	1404	GAGCAAGNACTGAGAAACTGGAGCTGGAGNTTGGCGGTGTAAAGAGATGATCAAGCT	1463	
Qy	401	GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu	420	
Db	1464	CAGCAGCAGCTCTTA---CAGCAGCTGGCCACCACGTTGTATGATGATGACACCCTCACTG	1520	
Qy	421	LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTrpThrLeuPhe	440	
Db	1521	TTTGGAGACTGTACTTGTCTGGAAGAAAGGAACGCTTAAAGAAAGAGTGGACCTTTT	1580	
Qy	441	LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg	460	
Db	1581	AAAGGCANAAAAAGAAATTTTGAGAGAGAAAGGCAAGCTTTACAGAAAGCTGCCATTCGA	1640	
Qy	461	LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnPhe	480	
Db	1641	TTGGGTTGGAGAGAAAGGCGTTTGAAGAAGAGCGCAGCTGGGTGAAAGCAGCAGTTT	1700	
Qy	481	LeuAsnMetThrAsnPheAspHisGlnAsnSerGlnAsnValLysLeuPheSerAlaPhe	500	
Db	1701	TTAAACATGACGAACCTTTGACCACCGAAGCTCAGAAAAATGTGAAACCTTTTTCAGTGCC	1760	
Qy	501	SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu	520	
Db	1761	TCAGGAAGTCTCGATCCAGACAAATCTTATAGTCCACTCAGGCCACGGCAAAAGAGCTA	1820	
Qy	521	HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla	540	
Db	1821	CACAGTGTGGCTAATGGGTGCCAGCTTGACATCAAACTGACTAAATCTCTTCTCTGCC	1880	
Qy	541	SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle	560	
Db	1881	TCACCTTCTACTTCAGACTTTTCGCCAGACACATTCATGTGTGTGTGTAACACAGTTCATC	1940	
Qy	561	SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr	580	
Db	1941	AGTGTGCTGAAATATACTCTGAAAGAAGGTAAACCAAGTAGGGTTGCAGAGAAAGCAGC	2000	
Qy	581	AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCys	600	

Db	2001	GATCAGAAGTGGAGCGTGCAGTCGAGCCCGAGCTCCGGGAGGGGTCTACAGCGGATGC	2066
Qy	601	<p> SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615 2061 TCTTCGGCCCTTCAGGAGCGCTCAGCGGGACCGAGATGACTTACCT 2105 </p>	
Db	2061	TCTTCGGCCCTTCAGGAGCGCTCAGCGGGACCGAGATGACTTACCT	2105
RESULT 4	AK077112	2245 bp mRNA linear HTC 02-SEP-2005	
LOCUS	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933406N16 product:HYPOTHETICAL 71.0 KDA PROTEIN homolog [Mus musculus], full insert sequence.		
ACCESSION	AK077112.1	GI:26097206	
VERSION	HTC; CAP trapper.		
KEYWORDS	Mus musculus (house mouse)		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.		
REFERENCE	1	Carninci, P. and Hayashizaki, Y.	
AUTHORS	High-efficiency full-length cDNA cloning		
TITLE	Meth. Enzymol. 303, 19-44 (1999)		
JOURNAL	10349636		
PUBLISHED	2		
REFERENCE	3		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
PUBLISHED	11042159		
REFERENCE	4		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
PUBLISHED	11076861		
REFERENCE	5		
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
PUBLISHED	6		
REFERENCE	7		
AUTHORS	The FANTOM Consortium, the RIKEN Genome Exploration Research Group Phase I and II Team.		
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
JOURNAL	Nature 420, 563-573 (2002)		
PUBLISHED	7		
REFERENCE	8		
AUTHORS	RIKEN Genome Exploration Research Group, Genome Science Group (Genome Network Core Team) and the FANTOM Consortium.		
TITLE	Antisense transcription in the Mammalian Transcriptome		
JOURNAL	Science 309, 1559-1563 (2005)		
PUBLISHED	8 (bases 1 to 2245)		
REFERENCE	9		
AUTHORS	The FANTOM Consortium, Riken Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group).		
TITLE	The Transcriptional Landscape of the Mammalian Genome		
JOURNAL	Science 309, 1559-1563 (2005)		
PUBLISHED	8 (bases 1 to 2245)		
REFERENCE	10		
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, S., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishikawa, K., Nomura, K., Numasaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Okumura, K., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		

Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tonaru, A., Toya, T., Yaeunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/

URL: http://fantom.gsc.riken.jp/.

FEATURES

source

1. .2245
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:4933406N16"
/db_xref="taxon:10090"
/clone="4933406N16"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"

misc_feature

1. .2245
/note="HYPOPHETICAL 71.0 KDA PROTEIN homolog [Mus musculus] (SPTR|AAH21749, evidence: FASTY, 99.7%ID, 100%length, match=1842)
putative"

ORIGIN

Alignment Scores:

Pred. No.: 4, 47e-303 Length: 2245
Score: 3140.50 Matches: 613
Percent Similarity: 99.7% Conservative: 0
Best Local Similarity: 99.7% Mismatches: 1
Query Match: 99.2% Indels: 1
DB: 6 Gaps: 1

US-10-644-084-2 (1-615) x AK077112 (1-2245)

QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
DB 325 ATGGGAGATTGGATGACTGTGACAGATCCAGTTCGTGTACAGAAAACAAAATCTCTCT 384
QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnValLeuCys 40
DB 385 CAATATACCTCAGAAACAAAGATGTCTCCGTCAGTTTGTACTCCAGCAAGTTCTGTGC 444
QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlu 60
DB 445 TCTTCAGTACCTTATCCAAAACAGTGCATGTTTCGGTGTCTTTCGACAGAGAG 504
QY 61 AsnileGluGlnSerileSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
DB 505 AACATTGAACAAAGATTTCTTATCTTGTATCAGGAGCTGACCACTTCGGTTTCCTCC 564
QY 81 LeuTyrGluGlnSerLysSerLysGluAlaLysArgGluLeuAsnIleValValLeu 100
DB 565 TTGTATGAAGAATCCAAAAGTAAAGAGGCAAGAGAGAAATTAATATAGTCGTCTCTG 624
QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
DB 625 AACTGTATGAACGAGCTGCTGTGTTTCAGCGGAAGAACCTCTGCTGGCCAGGAGCGGTG 684

QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
DB 685 GAGACACAGAACTTGAAGCTGGCAGTGACATGGACCCTCTGCAGAGCTGCTACGCCAAA 744
QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
DB 745 CTTAAGGAGCAGTTGGAAACGTCACGCGGGAGATGATCGGGCTTCAAGAGAGAGACAGG 804
QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
DB 805 CAGCTGCAGTGCACAGACAGGAGTTTGCATCAGCTCTCTGAGAGATGAGAAAGATGAGTTA 864
QY 181 GlnLysLeuGlnAsnIleIleIleSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
DB 865 CAAAAATTACAAAATATATCATAGCCGCGGTACTCAGTATATAATCATGTGTGAAGAGG 924
QY 201 LysGluAaGluTyrAsnLysLeuLysGluArgLysHisGlnLeuValMetAsnLysLys 220
DB 925 AAGGAGCGTGAATATAAATAAGCTAAAGAGCGCCCTGCATCAGCTCGTTATGAACAAGAAG 984
QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
DB 985 GATAAAACATAGCCATGATGTTTAAATTATGTGGTCTGAGCTGATGGCAACGAGGC 1044
QY 241 SerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
DB 1045 TCATCGAGGACTGACAAAACAGACGAGGAATGAAGATGAGATGTACAAAATCTCTGTTG 1104
QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysVal 280
DB 1105 AATGATTATGATACCGCCAGAGCAGATCCTGATGGAGAACCGCGAGCTGAAGAAGGTC 1164
QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
DB 1165 CTCTAGCAGATGAGAGAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1224
QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320
DB 1225 GAAAGAGCAGAGGACGCGCACAGGCACTGTTGCTATCTCCGATATAGAAGATGACTCTGG 1284
QY 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
DB 1285 GAACTGAGCAGACAGACGCGTGTGGGGCTTTCCTGTGACACTGTGAGAGAGCAGCTGACA 1344
QY 341 AsnSerIleAaGlyLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
DB 1345 AACACATCAGGAAACAGTGGAGAAATTTTGAAGAAGTCATGTAGAAAACTCGAATAACCA 1404
QY 361 AlaSerLysValHisSerGluGlyLeuAsnGluAspValIleSerArgGlnAspHis 380
DB 1405 GCTTCGAAGGTACACTCAGAGGGCTTAAATGAGGAGGAGCTCATCTCACGACAAGACCAT 1464
QY 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysGlyGluMetIleLysAla 400
DB 1465 GAGCAAGAGACTGAGAAAACGAGCTGGAGATTGAGCGGTGTAAAGAGATGATCAAGGCT 1524
QY 401 GlnGlnGlnLeuGlnGlnGlnLeuAlaThrCysAspAspThrThrSerLeu 420
DB 1525 CAGCAGCAGCTCTTA---CAGCAGCTGGCCACCACGTGTGATGATGACACCACTCCTG 1581
QY 421 LeuArgAspCysTyrLeuLeuGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440
DB 1582 TTTCGAGACTGTTACTTCTCTGCAAGAAAGAAACGCTTAAAGAGAGTGGACCTTTT 1641
QY 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaIleArg 460
DB 1642 AAAGCAAAAAAGAAATTTTGAAGAGAAAGGCGAAGCTTTTACAGAAGCTGCCATTCGA 1701
QY 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnPhe 480
DB 1702 TTGGGTTTGGAGAGAAAGCGCTTTTGAAGAGAGCGGCGCTGCTGGGTAAAGCAGCAGTTT 1761

QY	481	LeuAenMetThrAsnPheAspHisGlnAenSerGluAenValLysLeuPheSerAlaPhe	500
Db	1762	TTAAACATGACGAACTTTTGACCACACAGAACTCAGAAATGTGAAACTTTTCAGTGCCTTC	1821
QY	501	SerGlySerSerAspProAspHisLeuValHisSerArgProArgGlnLysLysLeu	520
Db	1822	TCAGGAAGTCTTGATCCAGACAACTCTTAGTCCACTCAGCGCCACGGCAAAAGAAGCTA	1881
QY	521	HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla	540
Db	1882	CACAGTGGCTGAATGGGTGCCAGCTTGACATCAAACTGACTAAATCTCTTCCTGCC	1941
QY	541	SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle	560
Db	1942	TCACCTTCTACTTCAGACTTTCGCCAGACACATTCATGTGTCTGACACAGATTCATC	2001
QY	561	SerValLeuAenIleThrProGluSerLysPheSerGluValAlaAargGluSerThr	580
Db	2002	AGTGTGCTGAATATACTCTCTGAAGAAAGTAAACCAAGTGAGGTTCGAAGAGAAACGACG	2061
QY	581	AepGlnLysTpsSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCys	600
Db	2062	GATCAGAAGTGGAGCGTGCAGTCGAGGCCACCTCGCGGAGGGGTGCTACAGCGGATGC	2121
QY	601	SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro	615
Db	2122	TCCTCGGCTTCAGGAGCGCTCACGGGACCGAGATGACTTACCT	2166
RESULT 5			
AK049080			
LOCUS			
DEFINITION	AK049080	2371 bp mRNA linear HTC 02-SEP-2005	
		Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length	
		enriched library, clone:C230097E14 product:hypothetical protein,	
		full insert sequence.	
ACCESSION	AK049080		
VERSION	AK049080.1	GI:26339745	
KEYWORDS		HTC; CAP trapper.	
SOURCE		Mus musculus (house mouse)	
ORGANISM		Mus musculus	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;	
		Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			

TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
COMMENT	
FEATURES	
source	
CDS	

Phase I and II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 RIKEN Genome Exploration Research Group, Genome Science Group (Genome Network Core Team) and the FANTOM Consortium.
Antisense Transcription in the Mammalian Transcriptome
Science 309, 1564-1566 (2005)

7 The FANTOM Consortium, Riken Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group).
The Transcriptional Landscape of the Mammalian Genome
Science 309, 1559-1563 (2005)

8 (bases 1 to 2371)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hasegaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayase, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gs.riken.jp).
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.
Location/Qualifiers
1. .2371
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:C230097E14"
/db_xref="taxon:10090"
/clone="C230097E14"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stages="0 day neonate"
437. .2290
/note="unnamed protein product; hypothetical protein (evidence: rscDS, ProCrest, decoder, NCBI CDS predictor, Longest-ORF)
putative"
/codon_start=1
/protein_id="BAC33536.1"
/db_xref="GI:26339746"
/translation="MGDMVTVDVLCETKNKLSQYTSRKMSPLSSYQQVLCSSVP LSKNVHGVFGVCTGNETQISYLDDELTTFGPSPSYEESKSEAKELNIVALNC MNELLVLQRKNLLAQESVETONLKLGDMDHLSQCYAKLKEQLTSREMLGQERDR KKDNIADVLNIVYGRADKGGKSWRTKTEARNEDEMYKILLNDEYKQKQLMENAE LKLYQLOKKEMISLLSPQKKKPRERAEADGTGTVAISDLSGELSVMGLSCDT VREQLTNSIRKQWRLKSHVKLDNQASKVHSEGLNEEDVTSRODHEQETKLEIE RKEMIKAQOQLLOLATTCDTDDTSLRLDCYLLLEERLKEEWTLPKEQKNFRER RSFTAAIRLGLERKAFEEERASWVKQFLNMTNFDHNSENVKLFSAFSGKPGSDP DNLIVHSRPROKHLHSVANGVACPCTSKLTSLPASPTSDFEROHCYSVSLVN ITPBESKPSPEVARESTQDKWSVQSRPSSREGICSGCSAFRSAGHRRDLP"

ORIGIN

Alignment Scores:

Pred. NO.:	8.68e-303	Length:	2371
Score:	3138.00	Matches:	614
Percent Similarity:	99.4%	Conservative:	0
Best Local Similarity:	99.4%	Mismatches:	0
Query Match:	99.1%	Indels:	4
DB:	6	Gaps:	2

US-10-644-084-2 (1-615) x AK049080 (1-2371)

Qy	1	MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer	20
Db	437	ATGGGAGATTGGATGACTGTGCACAGATCCAGTTCTGTGTACAGAAAAACAAAATCTCTCT	496
Qy	21	GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys	40
Db	497	CAATATACCTCAGAAACAAGATGTCCTCGTCCAGTTTGTACTCCAGCAAGATCTCTGGC	556
Qy	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu	60
Db	557	TCCTTCAGTACCTTTATCCAAAAACGTGTCATGGTGTTCGCTGTCCTTCGCACAGAGAG	616
Qy	61	AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer	80
Db	617	AACATTGAACAAGATTTCCTATCTTCATCAGGAGCTGACACCTCCGGGTTTCCTTCC	676
Qy	81	LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValalValLeu	100
Db	677	TTGTATGAAGAANTCCAAAGTAAAGAGGCAAGAGAGAAATTAATATATAGTCGCTGTTCTG	736
Qy	101	AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal	120
Db	737	AACGTATGAACGAGCTGCTCGTCTTCAGCGGAAGAACCTTCGTGGCCACAGAGAGCGTG	796
Qy	121	GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys	140
Db	797	GAGACACAGAACTTTGAAGCTGGCAGTCACATGGACCACTGCAGAGCTGTACGCCAAA	856
Qy	141	LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg	160
Db	857	CTTAAGGAGCAGATTGGAACAGCTCAGCGGGAGATGATCGGGCTTCAAGAGAGAGACAGG	916
Qy	161	GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal	180
Db	917	CAGCTGAGTGCAGAACAGAGGTTTGCATCAGCTCCTGAAGATGAGAAAGATGAGGTA	976
Qy	181	GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg	200
Db	977	CAAAAAATTAACAAAATATCATAGCCAGCGGGCTACTCAGTATATCATGATGTGAAGAGG	1036
Qy	201	LysGluArgGluTyrAsnLysLeuLysGluArgLysGlnLeuValMetAsnLysLys	220
Db	1037	AAGGAGCGTGAATATAATAAGCTAAAGGAGCGCTGCATCAGCTCGTTATGAACAAGAAG	1096
Qy	221	AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly	240
Db	1097	GATAAAAACAATAGCCATGGATGTTTTAAATTTATGTGGCTGAGCTGATGGCAACAGAGGC	1156
Qy	241	SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu	260
Db	1157	TCATGGAGGACTGCAAAAACAGAACCCAGGAATCAAGATGAGATGTACAAAATTCCTGTTG	1216
Qy	261	AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal	280
Db	1217	AATGATTATGAGTACCGCCGAACAGATCCTGATGGAGAACCGGGAGCTCAAGAGAGTC	1276
Qy	281	LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg	300
Db	1277	CTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTGCTCTCTCAGAGAAGAAGCCACGG	1336
Qy	301	GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly	320

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 1845)
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.

Inferring nonneutral evolution from human-chimp-mouse orthologous

gene trios

Science 302 (5652), 1960-1963 (2003)

JOURNAL

PUBMED

14671302

REFERENCE

AUTHORS

2 (bases 1 to 1845)

Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,

Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,

Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,

Adams, M.D. and Cargill, M.

Direct Submission

Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,

Rockville, MD 20850, USA

This sequence was made by sequencing genomic exons and ordering

them based on alignment.

Location/Qualifiers

1..1845

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

<1..>1845

/locus_tag="HCM4575"

ORIGIN

Alignment Scores:

Pred. No.: 18-267 Length: 1845

Score: 2786.50 Matches: 538

Percent Similarity: 93.2% Conservative: 35

Best Local Similarity: 87.5% Mismatches: 41

Query Match: 88.0% Indels: 1

DB: 14 Gaps: 1

US-10-644-084-2 (1-615) x AY412491 (1-1845)

Qy 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrsGluAsnLysAsnLeuSer 20

Db 1 ATGGGAGATTGGATGACTGTTACAGATCCAGTCTGTCTTCAGAAAGCAAACTATCTCT 60

Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40

Db 61 CAATATACCTCAGAAACCAAGATGCTCCATCAAGTTTATATCTACACCAAGTGTATGT 120

Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60

Db 121 TCTTCATACCTTTATCGAAAAATGTGCACAGTTTTTTTCAGTGCCTTTCTGCACACAGAT 180

Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80

Db 181 AATATTGAACACAGTATCTCATATCTTGATCAGGAATTGACTACTTTTGGTTTCTCTCA 240

Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100

Db 241 TTATATGAAGAAATCCAAAGGTAAAGAGACAAAGAGAGAGTTAAATATATAGTACTCTACTA 300

Qy 101 AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerVal 120

Db 301 AATTGTATGAATGAGCTGCTGCTGCTTCAGCCGGAAGACCTTCTAGCTCAGGAAAAATGTG 360

Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140

Db 361 GAGACACAGAAATTTGAAGCTGGGAAGTGATATGGACCATCTACAGAGCTGCTACTCAAAA 420

Qy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160

Db 421 CTTAAGGAACAACCTGGAAACCTCCAGGGGGAATGATTTGGGCTTCAGGAAAGACACAGA 480

Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGlnLysAspGluVal 180

Db 481 CAGTTACAAATGTAGAACAGGAATTTGCATCAGCTACTAAAGAAATGAGAAAGATGAGGTG 540

Qy 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200

Db 541 CAAAAATACAAAATATCATTCGAAGTCGAGCTACTAGTATAATCATCATATGAAGAGA 600

Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220

Db 601 AAAGAGCGTGAATATAATAAACTGAAGAACGCTCTACATCAACTTGTATATGAACAAAGAAA 660

Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240

Db 661 GATAAGAAAATAGCTATCGACATTTGAATATATGTCGGGAGAGCTGATGTAAGAAAGAGGC 720

Qy 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260

Db 721 TCCTGGAGGACTGGTAAAACTGAAGCCAGGAATGAAGATGAATGTATAAAATTTCTCTTG 780

Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280

Db 781 AATGATTATGAATATCGTCAGAAACAAATCTTAATGGAAAAATGCAAGAACTTAAGAAGGT 840

Qy 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300

Db 841 CTTCAACAAATGAAAAAGGAATGATTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900

Qy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320

Db 901 GAAAGAGTAGATGATAGTACAGGAACCTGTT---ATTTCCGATGTTGAAGAGAGATCCGGG 957

Qy 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340

Db 958 GAACCTAAGCAGAGAGATGATGGGACCTTTCTCTGTGAACATGTGAGAGAGCGCTTACA 1017

Qy 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360

Db 1018 AACGATCAGAAAACACTGGAGAAATTTGAAAAGTCATGTAGAAAAGCTTGTATAACCAA 1077

Qy 361 AlaSerLysValHisSerGluGlyLeuAsnGluAspValIleSerArgGlnAspHis 380

Db 1078 GTTTCAAAGGTACACCTGGAAGGTTTTAATGATGAAGATGTAATCTCAGCAAGACCAT 1137

Qy 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400

Db 1138 GAAACAAGAACTGAAAAAACTCGAGTTAGAAATTCAGCAGTGTAAAGAAATGATTAANAAT 1197

Qy 401 GlnGlnGlnLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420

Db 1198 CAGCAACAGCTTTTACACGACGCTCGCTACTGCATATGATGATGATACCACTTCACTA 1257

Qy 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440

Db 1258 TTACGAGACTGTTATTTGTTGGAAGAAAGAACGCTCTCAAAGAAAGATGGTCCCTTTT 1317

Qy 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460

Db 1318 AAAGACAGAAAAGAAATTTTGAAGGAGAGACGACGCTTTTACAGAGCCGCTATTCCG 1377

Qy 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPhe 480

Db 1378 CTGGGATTGGAGAAAGGCACTTTGAAGCAAGAAAGAGCCAGTTGGTTAAAGCAGCAGTTT 1437

Qy 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500

Db 1438 CTAATAATGACTACCTTTTGACCAACAGAACTCAGAAAAATGTGAAAATTTTCAGTGCCTTC 1497

Qy 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520

Db 1498 TCAGAAAGTCTGATTGGGACAACTTATATAGTCACTCGAGCAGCCGCAAGAAAGACCT 1557

Qy 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540

```

Db      1558 CACAGTGTCTAATGGTCTCCAGTTTGCATGTCCTAACTTACTAAATCTCTTCCTGCT 1617
QY      541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
Db      1618 TCACCTTCCACTTCAGACTTTTGCCAGACAGCGTTCCTGCATATCTCAACATAGTTCATC 1677
QY      561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580
Db      1678 AATGTACTGAATATAACTGCTGAAGAAATTTAAACCAATTCAGGTTCGAGGAGAAATGTACA 1737
QY      581 AspGlnLysThrSerValGlnSerArgProSerSerArgGluGluGlyCysTyrSerGlyCys 600
Db      1738 AATCAAAATGGAGTGTGCATCAAGACTTGGATCACAGGAGGTGCTATAGTGGATGC 1797
QY      601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
Db      1798 TCCTTGAGTACACAAATCTCATGTAGAAAAGATGACTTACCT 1842

RESULT 7
LOCUS   CR613292                2111 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0DL007Y108 of B cells (Ramos cell line)
          Cot 25-normalized of Homo sapiens (human).
ACCESSION CR613292
VERSION   1
KEYWORDS  HTC; CNSLT_cDNA.
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Hominoidea; Homo.
REFERENCE 1 (bases 1 to 2111)
AUTHORS   Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE     Full-length cDNA libraries and normalization
JOURNAL   Unpublished
REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
          http://fulllength.invitrogen.com/ invitrogen Corporation 1600
          Faraday Avenue
          2 (bases 1 to 2111)
          Genoscope.
REFERENCE 2 (bases 1 to 2111)
AUTHORS   Direct Submission
TITLE     Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL   BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
COMMENT   1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
          end enriched, double-strand cDNA was digested with Not I and cloned
          into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
          was normalized. Library was constructed by Life Technologies, a
          division of Invitrogen.
FEATURES
source   Location/Qualifiers
         1..2111
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DL007Y108"
            /tissue_type="B cells (Ramos cell line) Cot 25-normalized"
            /plasmid="pCMVSPORT_6"

ORIGIN
Alignment Scores:
Pred. No.:      1.24e-267      Length:      2111
Score:          2786.50      Matches:      538
Percent Similarity: 93.2%      Conservative: 35
Best Local Similarity: 87.5%      Mismatches:  41
Query Match:     88.0%      Indels:      1
DB:              6           Gaps:        1

US-10-644-084-2 (1-615) x CR613292 (1-2111)

QY      1 MetGlyAspTyrMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
Db      206 ATGGAGAGATTGGATGACTGTGTACAGATCCAGGTCTGTCTTCGAAAGCAAAACTATCTCT 265

```

```

QY      21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db      266 CAATATACCTCAGAACAAAGATGTCTCCATCAAGTTTATATCTACACAGCAAGTGTATGT 325
QY      41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db      326 TCTTCAATACCTTTATTCGAAAAATGTGCACAGTTTTTTTTCAGTGCCTTCTGCACAGAAGAT 385
QY      61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
Db      386 AATAATTGAACAGATATCTCATATCTTGATCAGGAATTCAGTACTATTTTGGTTTCTCTTCA 445
QY      81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db      446 TTATATGAAGAATCCAAAGGTAAAGAGACAAAGAGAGAGTAAATATATAGTAGCTGACTA 505
QY      101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
Db      506 AATTGTATGATGAGCTGCTTGTGCTTCAGCGGAAGAACCTTCTAGCTCAGGAAAAATGTG 565
QY      121 GluThrGlnAsnLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
Db      566 GAGACACAGAAATTTGAAGCTGGGAAGTGATATGGACCATCTACAGAGCTGCTACTCATAA 625
QY      141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
Db      626 CTTAAGGAACAACCTCGAAACCTCCAGGAGGGAATGATTGGGCTTCAGGAAAGAGACAGA 685
QY      161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180
Db      686 CAGTTTACAAATGTAAAGAACAGGAATTTGCATCAGCTACTTAAAGAATGAGAAGATGAGGTG 745
QY      181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Db      746 CAAAATTTACAAAATATCATTCAGAGTCGAGTACTCAGTATATCATGATATGAAGAGA 805
QY      201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db      806 AAAGAGCGTGAATATAATAAACTGAAGAACCGCTCTACATCAACTTGTATTGAACAAGAAA 865
QY      221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db      866 GATAAGAAAAATAGCTATGACATTTTGAATTTATATGTCCGAGAGAGCTGATGNAAGAGAGC 925
QY      241 SerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
Db      926 TCCTGGAGGAGCTGGTAAAACCTGAAGCCAGGAATGAAGATGAATGTATATAAATTTCTCTTG 985
QY      261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
Db      986 AATGATTATGAATATCGTCAGAAACAAATCCTTAATGGAAAAATGCAGAACTTAAAGAGGTT 1045
QY      281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
Db      1046 CTTCAACAAAATGAAAGAGAAATGATTTCTTCTTCTTCTCCCAAAAGAAAGAAACCTAGA 1105
QY      301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
Db      1106 GAAAGAGTAGATGATAGTACAGGAACCTGTT---ATTTCCGATGTTGAGAAAGATGCCGGG 1162
QY      321 GluLeuSerArgAspSerValTyrGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
Db      1163 GAACTTAAGCAGAGAGAGATGTGGGACCTTCTCTGTGAAACTGTGAGAGAGCAGCTTACA 1222
QY      341 AsnSerIleArgLysGlnTyrArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
Db      1223 AACAGCATCAGAAAACAGTGGAGAAATTTTGAAGAGTCTGTAGAAAAGCTGTGATAACCAA 1282
QY      361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
Db      1283 GTTTCAAAAGGTACACCTCGAAGGTTTTTAATGATGAAGATGTAATCTCAGCAGACAGACCAT 1342

```

```

Qy 381 GluGlnGluThrGluLysLeuGluLeuGluLeuGluArgCysGluMetileLysAla 400
Db 1343 GAACAGAAACTGMAAACTCAGTTAGAAATTCAGCAGTGTAAAGAAATGATTAACACT 1402
Qy 401 GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspThrThrSerLeu 420
Db 1403 CAGCAACAGCTTTTACAGCAGCAGCTCGTACTGCTATGATGATGATACCACTTCACTA 1462
Qy 421 LeuArgAspCysTyLeuLeuGluGluGluArgGluArgLeuLeuGluThrThrLeuPhe 440
Db 1463 TTACGAGACTGTATTGTTGTAAGAAAGAAAGAACGCTCTCAAAAGAAAGATGGTCCCTTTT 1522
Qy 441 LysGluGlnLysLysAsnPheGluArgGluArgGluArgSerPheThrGluAlaAlaLeuArg 460
Db 1523 AAAGAGCAGAAAGAAAGATTTTGAGAGGGAGAGACGAGCTTTTACAGAGCCGCTATTGCG 1582
Qy 461 LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnGlnPhe 480
Db 1583 CTGGGATTGGAGAGAAAGCAATTTGAAGAAGAAAGAGCAGCTTGGTTAAAGCAGCAGTTT 1642
Qy 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
Db 1643 CTAAATATGACTACCTTTTGACCAACAGAACTCAGAAATGTGTAAGCAGCAGTTT 1702
Qy 501 SerGlySerSerAspProAspAsnLeuLeuValHisSerArgProArgGlnLysLysLeu 520
Db 1703 TCAGAAAGTTCTGATTGGACAACTTTATATGTCGACTCGAGCAGCCGCAAAAGAGCCT 1762
Qy 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db 1763 CACAGTGTCTAAATGGGTCTCCAGTTGTCATGCTCTAAACTTACTAAATCTCTCTCTGCT 1822
Qy 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerLeu 560
Db 1823 TCACCTTCACCTTCAGACTTTTTCACAGACAGCTTCTCGATATCGAACATAGTTCAATC 1882
Qy 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgLysThr 580
Db 1883 AATGACTGAAATTAAGTCTGAGGAAATTAACCAATCAGTTGGAGGAGAAATGTACA 1942
Qy 581 AspGlnLysTrpSerValGlnSerArgProSerArgGluGluGlyCysTrpSerGlyCys 600
Db 1943 AATCAAAATAGGAGTGTGCATCAAGACCTGGATCACAGAGAGTTGCTATAGTGGATGC 2002
Qy 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
Db 2003 TCCTTGAGCTACACAAATTTCTATGTAGAAAAGATGACTTACCT 2047

RESULT 8
LOCUS BC035580
DEFINITION Homo sapiens synovial sarcoma, X breakpoint 2 interacting protein,
mRNA (CDNA clone IMAGE:5575710).
ACCESSION BC035580
VERSION BC035580.1 GI:22028261
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 3154)
AUTHORS Strausberg,R.D., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.P., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Uedin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Rana,S.S., Loquellano,N.A., Peters,G.J.,
Armstrong,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,

```

```

Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywicki,M.I., Skalska,U., Smalish,D.E.,
Schnorch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Mammalian Gene Collection Program Team
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 3154)
.
NIH MGC Project
Direct Submission
Submitted (31-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@hgrl.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.B., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Skatripp,S., Thomas,P.J., Touchman,J.W.,
Turgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 69 Row: n Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 51317360
This clone has the following problem: no 5' EST match.
FEATURES             Location/Qualifiers
     source            1..3154
     -organism="Homo sapiens"
     -mol_type="mRNA"
     -db_xref="taxon:9606"
     -clone="IMAGE:5575710"
     -tissue_type="Testis, embryonal carcinoma"
     -clone_lib="NIH_MGC_92"
     -lab_host="DH10B"
     -note="vector: pCMV-SPORT6"
ORIGIN
Alignment Scores:      2,33e-267      Length:      3154
Pred. No.:            2786,50      Matches:      538
Score:                93.2%      Conservative: 35
Percent Similarity:   93.2%      Mismatches:   41
Best Local Similarity: 87.5%      Indels:       1
Query Match:         88.0%      Gaps:         1
Db:
US-10-644-084-2 (1-615) x BC035580 (1-3154)
Qy 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysLeuSer 20
Db 86 ATGGAGATTGGATGACGTGTTACAGATCCAGCTGCTCTTCAGAAAGCAAACTATCTCT 145
Qy 21 GlnTyThrSerGluThrLysMetSerProSerSerLeuTySerGlnGlnValLeuCys 40
Db 146 CAATATACCTCAGAAACAAGATGTCTCCATCAAGTTTATATCTCAGCAAGTGTATGT 205

```

41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
 206 TCITCAATACCTTTATCGAAAAATGTGCACAGTTTTTTTCAGTGCCTTCTGCACAGAAAT 265
 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGlnLeuThrThrPheGlyPheProSer 80
 266 AATATTGAACAGAGTATCTCATATCTTGATCAGGAATTGACTATTGGTTTCTCTCA 325
 81 LeuTyrGluGlnSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
 326 TTATATCAAGAAATCCAAAGGTAAAGACACAAAGAGAGAGTTAAATATAGTAGCTACTA 385
 101 AsnCysMetAsnGlnLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerVal 120
 386 AATTGTATGAATCAGCTGCTGTGTCTTCAGCGAAGAACCTTCTAGCTCAGSAAATGTG 445
 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
 446 GAGACACAGAAATTTGAAGCTGGAGAGTGATATGGACCATCTACAGAGCTGCTACTCAAA 505
 141 LeuLysGluGlnLeuGluThrSerArgGluMetIleGlyLeuGlnGluArgAspArg 160
 506 CTTAAGGAAACAACTGGAACCTCCAGAGGGAATGATTGGGCTTCAGGAAACAGACAGA 565
 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
 566 CAGTTACAATGTAAAGAACAGGAATTTGCATCAGCTACTAAAGAAATGAGAAGATGAGGTG 625
 181 GlnLysLeuGlnAsnIleLeuAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
 626 CAAAATTAACAANAATATCATTTGCAAGTCGAGCTACTCAGTATATCATGATATGAAGAGA 685
 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
 686 AAAGAGCGTGAATATATAAATGGAAGAACGCTCATCACTCACTGTTATGAACAAGAAA 745
 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
 746 GATAAGAAATAGCTATGACATTTTGAATTTATGTCTGGAGAGCTCATGGAAGAAAGAGGC 805
 241 SerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
 806 TCTCGAGGAGCTGGTAAACTGAAGCCAGGAATGAAGATGAATATATATAAATCTCTTG 865
 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
 866 AATGATTATGAATATCGTCAGAAAACAAATCCCTAAATGGAAAATGCAGAACTTAAGAGGTT 925
 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
 926 CTTCAACAAATGAAAAGGAAATGATTTCTCTTCTTCTCCCAAAAGAAAGAACCTAGA 985
 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
 986 GAAGAGTAGATAGTAGTACAGGAAGTGT---ATTTCGATGTTGAAGAGATGCCGGG 1042
 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
 1043 GAACTAAGCAGAGAGAGTATGTGGGACCTTTCTGTGAAACTGTGAGAGAGCAGCTTACA 1102
 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
 1103 AACAGCATCAGAAAACAGTGGAGAAATTTTGAAGAGTCATGTAGAAAAGCTTGATACCAA 1162
 361 AlaSerLysValHisSerGluGlyLeuAsnGluAspValIleSerArgGlnAspHis 380
 1163 GTTTCAAAGGTACACCTCGAAGGTTTTAATGATGAAGATGATATCTCAGCAAGACCAT 1222
 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400
 1223 GAACAAGAAACTGAAAACTCGAGTTAGAAAATTCAGCAGTGTAAAGAAATGATTAACACT 1282

401 GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
 1283 CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCATATGATGATATACCACTTCACTA 1342
 421 LeuArgAspCysTyrLeuLeuGluLysGluArgLeuLysGluGluTyrThrLeuPhe 440
 1343 TTACGAGACTGTTATTTTGTGGAAAGAAAGAACGCTCTCAAGAGAAATGGTCCCTTTT 1402
 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
 1403 AAAGAGCAGAAAAAGAAATTTTGAGAGGGAGAGACGAAGCTTTACAGAAGCGCTATTCCG 1462
 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPhe 480
 1463 CTGGGATTGGAGAGAAAGGCATTTGAAGAGAAAGAGCCAGCTTGGTTAAAGCAGCAGT 1522
 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
 1523 CTAAATATGACTACCTTTGACCAACCAAGAACTCAGAAATGTGAAACTTTTCAGTGCCTC 1582
 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
 1583 TCAGGAAGTCTGATTGGGACAATCTTATAGTCACCTCGAGGCGAGCGCAAAAGAGCCCT 1642
 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
 1643 CACAGTGTCTAAATGGGCTCTCCAGTTTGCATGTCTTAAACTTACTAAATCTCTTCCCTG 1702
 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
 1703 TCACCTTCCACTTCAGACTTTTGGCCAGACACGTTCTCTGCATATCTGAACATAGTTCAATC 1762
 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580
 1763 AATGTACTCAATATAACTCTCTGAAGAAATTTAAACCAATATCAGGTTGGAGAGAAATGTACA 1822
 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluClyCysTyrSerGlyCys 600
 1823 AATCAAAAATGGAGTGGCGTCAAGACCTGGATCATCACGAAGGTTGCTATAGTGGATGC 1882
 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615
 1883 TCTTTCAGCTACACAAATTTCTCATGTAGAAAAGATGACTTACCT 1927
 RESULT 9
 HSM801317
 LOCUS Homo sapiens mRNA; cDNA DKFZp434C0515 (from clone DKFZp434C0515).
 DEFINITION
 ACCESSION AL133046
 VERSION AL133046.1 GI:6453458
 KEYWORDS HTC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 5658)
 AUTHORS Bloecker,H., Boecker,M., Brandt,P., Mewes,H.W., Weil,B., Amid,C.,
 Osanger,A., Fobo,G., Han,M. and Wiemann,S.
 CONSRPTM The German cDNA Consortium
 JOURNAL Direct Submission
 COMMENT Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by GBF (National Research Centre for Biotechnology Ltd.,
 Braunschweig/Germany) within the cDNA sequencing consortium of the
 German Genome Project.
 This clone (DKFZp434C0515) is available at the RZPD Deutsches
 Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
 Please contact RZPD for ordering:
 http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp434C0515
 Further information about the clone and the sequencing project is

available at <http://mips.gsf.de/projects/cdna/>.

FEATURES		Location/Qualifiers	
source	1..5658	/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="RZPD:DKFZp434C0515"	
		/db_xref="taxon:9606"	
		/clone="DKFZp434C0515"	
		/issue_type="testis"	
		/clone_lib="434 (synonym: htes3). Vector pSport1; host DH10B; sites NotI + SalI"	
		/dev_stage="adult"	
		/note="synovial sarcoma, X breakpoint 2 interacting protein"	
		1..5658	
gene		/gene="DKFZp434C0515"	
CDS	267..2111	/gene="DKFZp434C0515"	
		/codon_start=1	
		/product="hypothetical protein"	
		/protein_id="CAB61373.1"	
		/db_xref="GI:6453459"	
		/db_xref="UniProtKB/TrEMBL:Q9UIX0"	
		/translation="MGDMWTVDPLGSSEKTIISQYTSFTKMSPSLSYQQVLCSSIP LSKNVHSPFSAFCTEDNTQISYLDQSLTTFPPSLYVEESKGKTKRELNIIVALNC MNELAVLQKNLADENVETONLKLGSMDHLQSCYSKLKQLETSRREMIGLQERDR QLOCKNRNLHOLLKNEDEVKQLNLIASRATQYNHDMKREYENKLRHLQLVN KKKIAMDILNVRAGDQKSWRTGKTEARNEDEMYKILLNDYFKQKLMENAE LKKVLOQMKEMIISLSPQKKPRERVDSTLVSDEEDAGELSRSDMDLSCETV REQLTNSRKQWRILKSHVEKLNDQVSKVHLEGFNDEDVI SRQHEQETKLELEIQ KEMTKTQLOQLATAYDDTSLRLDQLEEKRLKEWMLFKEQKKPFRER RSPTAAJRLGLERKA FREERASWLKQOFLANWTFDHOHNSENVKLFSAFSGSSDNDL LVHSQPOKKHSHVSNVSGPVCMSKLTSLPASPSDSFCQTRSCISEHSSINVLNITA BEIKPNQGGERTNOKWSVASRPGSQEGCYSGCSLSYSTNSHVEKDDLP"	
		ORIGIN	
		Alignment Scores:	
	Pred. No.:	5,866-267	Length: 5658
	Score:	2786.50	Matches: 538
	Percent Similarity:	93.2%	Conservative: 35
	Best Local Similarity:	87.5%	Mismatches: 41
	Query Match:	88.0%	Indels: 1
	DB:	6	Gaps: 1
US-10-644-084-2 (1-615) x HSM801317 (1-5658)			
Qy	1	MetGlyAspTyrMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20	
Db	267	ATGGGAGATTGGATGACTGTTACAGATCCAGGCTGCTCTTCAGAAAGCAAACTATCTCT 326	
Qy	21	GlnTyrThrSerGluThrIlyeMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40	
Db	327	CAATATACCTTCAGAAACCAAGATGCTCCATCAAGTTTATATCTACAGCAAGTGTATGT 386	
Qy	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60	
Db	387	TCCTCAATACCTTTATCGAAATATGTCACAGATTTTTCAGTGCCTCTGTCACAGAGAT 446	
Qy	61	AsnIleGluGlnSerIleSerTyrLeuAspGlnGlnLeuThrThrPheGlyPheProSer 80	
Db	447	AAATATTGAACAGAGTATCTCATATCTTGATCAGGAATTGACTTTTGGTTTCTCTTCA 506	
Qy	81	LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100	
Db	507	TTATATGAAGAAATCCAAAGGTAAAGAGACAAAGAGAGAGATTAAATATATAGTACTACTA 566	
Qy	101	AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120	
Db	567	AAATTGTATGAATGAGCTGCTTGTGCTTCAGCGGAGAGACCTTCTAGCTCAGGAAATGTG 626	
Qy	121	GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140	
Db	627	GAGACACAGAAATTTGAAGCTGGGAAATGATATGGACCCTCTACAGAGCTGTCTACTCAAAA 686	

Qy	141	LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
Db	687	CTTAAGGAACAACCTGGAAACCTCCAGGAGGAGAAATGATTGGCTTCAGGAAGACACAGA 746
Qy	161	GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
Db	747	CAGTTACAAATGTAGAACAGCAATTTGCATCAGCTACTAAAGATGAGAAAGATGAGGTG 806
Qy	181	GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Db	807	CAAAAAATACAAAATATCATTCGAAGTCGAGCTACTCAGTATAATCATGATATGAAGAGA 866
Qy	201	LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db	867	AAAGAGCGTGAATATAATAAAGCTGAAGAAAGCTCATCATCACTTGTATTGAACAGAAA 926
Qy	221	AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db	927	GATRAAGAAAATAGCTATCGACATTTTGAATTTATGTCGGGAGAGCTGATGGAAGAGGC 986
Qy	241	SerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrIlyeIleLeuLeu 260
Db	987	TCCTGGAGGAGCTGGTAAACTGAAGCCAGGAATGAAGATGAATAATGATAAAATTTCTCTG 1046
Qy	261	AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
Db	1047	ATGATTTATGAATATCGTCAGAAACAAATCTTAATGGAAATGCAAGACTTAAGAAGGTT 1106
Qy	281	LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
Db	1107	CTTCAACAAATGAAAGAGAAATGATTTCTCTCTTTCTCCCAAAAGAGAAACCTACA 1166
Qy	301	GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320
Db	1167	GAAGAGTAGATGATAGTACAGGAAGTCTTATTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1223
Qy	321	GluLeuSerArgAspSerValTyrGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
Db	1224	GAACTAACAGAGAGAGATGATGGGACCTTTCTCTGTGAACCTGTGAGAGAGAGAGCTTACA 1283
Qy	341	AsnSerIleArgLysGlnTyrArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
Db	1284	AACAGCATCAGAAACACAGTGGAGAAATTTTGAAGATCATGTAGAAAAGCTTGATAACCAA 1343
Qy	361	AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
Db	1344	GTTTCAAGAGGTACACCTCGAAGGTTTAAATGATGAAGATGTAATCTCAGCAAGACCAT 1403
Qy	381	GluGlnGluThrGluLysLeuGluLeuGluArgCysLysGluMetIleLysAla 400
Db	1404	GAACAGAAACTGAAAAACTCGAGTTAGAAATTCAGCAGTGTAAAGAAATGATTAAGACT 1463
Qy	401	GlnGlnGlnLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
Db	1464	CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGTCATATGATGATGATGATGATGATGAT 1523
Qy	421	LeuArgAspCysTyrLeuLeuGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440
Db	1524	TTACGAGACTGTTATTTGTTGGAAGAAAGGAAAGCTCTCAAGAAAGAAATGGTCCCTTTT 1583
Qy	441	LysGluGlnLysLysAsnPheGluArgGluArgArgSerPheThrGluAlaAlaIleArg 460
Db	1584	AAAGAGCAGAAAGAAAGAAATTTTGAAGGAGAGAGAGACGAGCTTTACAGAGCCGCTATTCCG 1643
Qy	461	LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTyrPValLysGlnGlnPhe 480
Db	1644	CTGGCATTTGGAGAGAGAGGCAATTTGAAGAAAGAGAGGCGAGCTGGTTGTTAAAGCAGCAGTTT 1703
Qy	481	LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
Db	1704	CTAAATATGACTACTCTTTGACCCACAGAACTCAGAAATATGTGAAACTTTTTCAGTGCCTTC 1763

QY 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
 Db 1764 TCAGGAAGTCTCATTTGGGACAATCTTATAGTCACCTCGAGGCGCGCAAAAAGACCT 1823
 QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
 Db 1824 CACAGTGTGCTAATGGGTCTCCAGTTTGATGCTCTAAACTTACTAAATCTCTCTGCT 1883
 QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerile 560
 Db 1884 TCACCTTCCACTTCAGACTTTTCCAGACACGTTCTCTGCATATCTGAACATAGTCAATC 1943
 QY 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580
 Db 1944 AATGTACTGAATTAATCTGCTGAAGAAATTAACCAAAATCAGGTGGAGAGAACGTACA 2003
 QY 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGlyCys 600
 Db 2004 AATCAAAAATGGAGTGTGGCGTCAAGACCTGGATCACAGGAAGTTGCTATAGTGGATGC 2063
 QY 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615
 Db 2064 TCCTTGAGCTACAAAATTCATCTAGTAAAAAGATCACTTACCT 2108

RESULT 10

AY412492

LOCUS

DEFINITION Pan troglodytes HCM4575 gene, VIRTUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.

ACCESSION AY412492

VERSION AY412492.1

KEYWORDS GI:39768457

SOURCE Pan troglodytes (chimpanzee)

ORGANISM

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominoidea; Pan.

1 (bases 1 to 1845)

Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,

Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,

Fertig, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,

Adams, M.D. and Cargill, M.

Inferring nonneutral evolution from human-chimp-mouse orthologous

gene trios

Science 302 (5652), 1960-1963 (2003)

14671302

2 (bases 1 to 1845)

Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,

Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,

Fertig, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,

Adams, M.D. and Cargill, M.

Direct Submission

Submitted (16-NOV-2003)

Rockville, MD 20850, USA

This sequence was made by sequencing genomic exons and ordering

them based on alignment.

Location/Qualifiers

1..1845

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

<1..1845

/locus_tag="HCM4575"

ORIGIN

Alignment Scores:

Pred. No.: 8,13e-264 Length: 1845

Score: 2747.50 Matches: 531

Percent Similarity: 92.0% Conservative: 35

Best Local Similarity: 86.3% Mismatches: 48

Query Match: 86.8% Indels: 1

DB: 14 Gaps: 1

gene

source

1..1845

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

<1..1845

/locus_tag="HCM4575"

ORIGIN

US-10-644-084-2 (1-615) x AY412492 (1-1845)
 QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
 Db 1 ATGGGAGATTGGATGACTGTTACAGATCCAGGCTCTCTTCAGAAAGCAAACTATCTCT 60
 QY 21 GlnTyThrSerGluThrLysMetSerProSerSerLeuTySerGlnGlnValLeuCys 40
 Db 61 CAATATACCTTCAGAAACAAAGATGTCCTCATCAAGTTNNNACTCACAGCAAGTGTATGT 120
 QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
 Db 121 TCTTCAATACCTTTATCGAAATAATGTGCACAGTTTTTTCAGTCCCTCTTCGCACAGAGAT 180
 QY 61 AsnIleGluGlnSerIleSerTyLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
 Db 181 AATATTGAACACAGATATCTCATATCTTGATCAGGAATTCAGTACTTTTGGTTTTCTTCA 240
 QY 81 LeuTyGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
 Db 241 TTATATGAAGAAATCCAAAGGTAAAGAGACAAAGAGAGAGTTAAATATAGTAGTGCTACTA 300
 QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
 Db 301 AATTGTATGAATGAGCTGCTTGTGCTTCACGGGAAGAACCTTCTAGCTCAGGAAAATGTG 360
 QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyAlaLys 140
 Db 361 GAGACACAGAAATTTGAAGCTGGGAAGTGATATGGACCATCTACAGAGCTGCTACTCAAAA 420
 QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
 Db 421 CTTAAGGAACAACTGGAAACCTCCAGGAGGGAATGATTGGGCTTCAGGAAAGAGACAGA 480
 QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180
 Db 481 CAGTTACAAATGTAAAGAACACAGGAATTTGCACTACAGTACTAAAGAAATCAGAAAGATGGTG 540
 QY 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyAsnHisAspValLysArg 200
 Db 541 CAAAAATTTACAAAAATATCATTGCAAGTCGAGCTACTCAGTATATATCATGATGNNNNNN 600
 QY 201 LysGluArgGluTyAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
 Db 601 AAAGAGCGTGAATATAATAANNNGNNNNAACGCTCTACATCAACTTCTTATGAACAGAAA 660
 QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyValGlyValArgAlaAspGlyLysArgGly 240
 Db 661 GATAAGAAATAGCTATGGACATTTTGAATTTATGTCGGGAGAGCTGATGGAAGAGAGGC 720
 QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyLysIleLeuLeu 260
 Db 721 TCCTGGAGGACTGGTAAACCTCAAGCCAGGAATGAAGATGAATGTATATAAATTCCTTTG 780
 QY 261 AsnAspTyArgGluTyArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysVal 280
 Db 781 AATGATTATGAATATCGTCAGAAACAAATCCTTAATGGAAAAATGCAGAACTTAAGAGGTT 840
 QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
 Db 841 CTTCAACAAATGAAAAAGAAATGATTTCTCTCTTCTTCTCCCAAAAAGAGAAACCTTAGA 900
 QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320
 Db 901 GAAAGAGTAGATGATAGTACAGGAATCTGTT---ATTCTGATGTTGAGAAAGATGCCGGG 957
 QY 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
 Db 958 GAACTAAGCAGACAGAGATGATGGGACCTTTCTCTGTAACCTGTGAGAGAGCAGCTTACA 1017
 QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
 Db 1018 AACAGCATCAGAAAAACAGTGGAGAAATTTTGAANAAGTCATGTAGAAAAAATCTTATAACCAA 1077

```
Qy 361 AlaSerLysValHisSerGluGlyLeuAenGluGluAaspValHisSerArgGlnAaspHis 380
Db 1078 GTTTCAGAGGTACACCTCGAAGGTTTATGATGATGAAGATGTTATCTCAGCACAAGACCAT 1137
Qy 381 GluGlnGluThrGluLysLeuGluLeuGluLysGluAArgCysGlyGluMetIleLysAla 400
Db 1138 GAACAAGAACTGMAAACTCGAGTTAGAAATTCAGCAGTGTAAAGAAATGATTAAGACT 1197
Qy 401 GlnGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 420
Db 1198 CAGCAACAATTTTACAGCAGCAGCTCGCTACTGCTATGATGATGATGATGATGATGATGAT 1257
Qy 421 LeuArgAepCysTyrLeuLeuGluGluLysGluAArgGlnGlnGlnGlnGlnGlnGlnGln 440
Db 1258 TTACGAGACTGTTATTTGTTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1317
Qy 441 LysGluGlnLysLysAenPheGluAArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 460
Db 1318 AAAGAGCAGAAAAAGAAATTTTTCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
Qy 461 LeuGlyLeuGluArgLysAlaPheGluGluGluAArgGlnGlnGlnGlnGlnGlnGlnGln 480
Db 1378 CTGGGATTCGAGAGAAAGCAATTTGAAGAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAG 1437
Qy 481 LeuAenMetThrAenPheAaspHisGlnAenSerGluAenValLysLeuPheSerAlaPhe 500
Db 1438 TTAAATATGACTACCTTTGACCACACAGAACTCAGAAATGCGAAATTTTCAGTCCCTTC 1497
Qy 501 SerGlySerSerAaspPheAenLeuValHisSerArgProAArgGlnLysLysLeu 520
Db 1498 TCAGGAAGTCTGATTTGGGACAATCTTATAGTGCACTCGAGGAGAGAGAGAGAGAGAG 1557
Qy 521 HisSerValAlaAenGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db 1558 CACAGTGTCTTAATGGGCTCCAGTTTGCATGCTTAAACTTACTTAAATCTCTTCTGCT 1617
Qy 541 SerProSerThrSerAaspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
Db 1618 TCACCTTCACCTTCAGACTTTTCCAGACACAGCTTCTCGCATGCTCGAACATGATTCATC 1677
Qy 561 SerValLeuAenIleThrProGluGluSerLysProSerGluValAlaAArgGluSerThr 580
Db 1678 AATGACTCAATATACTGCTGAAGAAATTAACCAAAATCAGGTTGGAGGAGAAATGTACA 1737
Qy 581 AspGlnLysThrSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCys 600
Db 1738 AATCAAAAATGGAGTGTGGCGTCAAGACCTGGATCACAGGAAGGTTGCTATAGTGGATGC 1797
Qy 601 SerSerAlaPheArgSerAlaHisGlyAepArgAaspAaspLeuPro 615
Db 1798 TCCTTGAGCTACACAAATTCATGTAGAGAAAAGATGACTTACCT 1842
```

```
RESULT 11
CX565842
LOCUS
DEFINITION
  UI-M-IB0-cum-q-09-0-UI-r1 NIH BMAP_IB0 Mus musculus cDNA clone
IMAGE:30945752 5', mRNA sequence.
ACCESSION
  CX565842
VERSION
  CX565842.1 GI:57592871
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 878)
AUTHORS
  NIH-MGC http://mgc.ncl.nih.gov/.
TITLE
  National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
  Unpublished (1999)
COMMENT
  Contact: Robert Strausberg, Ph.D.
  Email: cgpbbs-r@mail.nih.gov
  Tissue Procurement: Dr. James Lin University of Iowa
```

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

FEATURES

source

Location/Qualifiers

1..878
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30945752"
/tissue_type="whole eye"
/dev_stages="newborn (1,5,15 days)"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP IB0"
/note="Organ: Eye; Vector: pYX-Ase; Site 1: Ecor I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pYX-Ase vector. The library tag
sequence located between the Not I site and the polyA tail
is AATAATTACG. This library was created for the University
of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:

Pred. No.: 1..86-137 Length: 878
Score: 1482.00 Matches: 291
Percent Similarity: 99.7% Conservative: 0
Best Local Similarity: 99.7% Mismatches: 1
Query Match: 46.8% Indels: 0
DB: 9 Gaps: 0

US-10-644-084-2 (1-615) x CX565842 (1-878)

```
Qy 104 AenGluLeuLeuValLeuGlnArgLysAenLeuLeuAlaGlnGluSerValGluThrGln 123
Db 2 AACGAGCTGCTGCTGCTTCAGCGGAAGAACTGCTGGCCCCAGGAGAGCGTGGAGACAG 61
Qy 124 AenLeuLysLeuGlySerAaspMetAaspHisLeuGlnSerCysTyrAlaLysLeuLysGlu 143
Db 62 AACTTGAAGCTGGGCGAGTGACATGGACCACCTGCGAGAGCTGCTACGCCCAACTTAAGCAG 121
Qy 144 GlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGlnGluArgAaspArgGlnLeuGln 163
Db 122 CAGTTGGAACCTCCAGCGGAGATGATCGGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 181
Qy 164 CysLysAenArgSerLeuHisGlnLeuLysAenGluLysAenGluValGlnLysLeu 183
Db 182 TGCAAGAACAGGAGTTTGCATCAGCTCCTGAAGAATGAGAAGATGAGGATGACAAAATTA 241
Qy 184 GlnAenIleAlaSerArgAlaThrGlnTyrAenHisAaspValLysAargLysGluArg 203
Db 242 CAAAATATCATGACCGCGGCTACTCAGTATATCATGATGTGAAGAGAGAGAGAGCGT 301
Qy 204 GluTyrAenLysLeuLysGluArgLeuHisGlnLeuValMetAenLysLysAaspLysAen 223
Db 302 GAATATAATAGCTAAAGAGAGCGCTGCATCAGCTCGTTATGACACAGAGAGATAAAAAC 361
Qy 224 IleAlaMetAaspValLeuAenTyrValGlyArgAlaAaspGlyLysAtrGlySerTrpArg 243
Db 362 ATAGCCATGGATGTTTAAATATGATGCGGTGAGCTGATGCGCAACAGGCTCATGGAGG 421
```


QY 358 AspAsnGlnAlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArg 377
 |||||
 Db 840 GATAACCAAGTTTCAAAAGTACACTTGAAGGTTTTTAACGATGAGATGATATCTCACGA 899

QY 378 Gln 378
 ||||
 Db 900 CAA 902

RESULT 13
 CN539511 804 bp mRNA linear EST 29-APR-2004
 LOCUS UI-M-HUO-cgu-n-01-0-UI.r1 NIH_BMAP_HUO Mus musculus cDNA clone
 DEFINITION IMAGE:30667224 5', mRNA sequence.

ACCESSION CN539511

VERSION CN539511.1 GI:46867667

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 804)

REFERENCE NIH-MGC <http://mgs.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. James Lin University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

<http://genome.uiowa.edu/distribution/mousefl.html>

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1..804

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:30667224"

/tissue_type="whole eye"

/dev_stage="newborn(1, 5, 15 days)"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH_BMAP_HUO"

/note="Organ: Eye; Vector: pYX-Asc; Site:1: EcoR I;

Site 2: Not I; The library was constructed according

Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. Denatured RNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with oligo-dT

primer containing a Not I site. Double strand cDNA was

size selected according to mRNA size fraction, ligated

with EcoR I adaptor, digested with NotI and then cloned

directionally into pYX-Asc vector. The library tag

sequence located between the Not I site and the polyA tail

is AATATTAG. This library was created for the University

Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the

Developing Mouse Nervous System', supported by National

Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:

Pred. No.: 1,05e-125 Length: 804

Score: 1364.00 Matches: 265

Percent Similarity: 99.3% Conservative: 0

Best Local Similarity: 99.3% Mismatches: 2

Query Match: 43.1% Indels: 0

DB: 8 Gaps: 0

US-10-644-084-2 (1-615) x CN539511 (1-804)

QY 190 ArgAlaThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLys 209
 |||||
 Db 2 CGGCTACTCTAGTATAATCATGATGTGAAGAGGAGCGGTGAATAATAAAGCTAAAG 61

QY 210 GluArgLeuHisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeu 229
 |||||
 Db 62 GAGCGCTGTCATCAGCTCGTTATGAACAGAGGATAAAAACATAGCCATGATGTTTTTA 121

QY 230 AsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAla 249
 |||||
 Db 122 AATTATGTGGGTGCGAGCTGTGGCAACAGAGCTCATGGAGACTTGACAAAACAGAAGCC 181

QY 250 ArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGln 269
 |||||
 Db 182 AGGAATGAAGATGAGTGTACAAAATTCTGTTGAATGATTATAGTACCGCCAGAAGCAG 241

QY 270 IleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIle 289
 |||||
 Db 242 ATCTGTATGGAGAACGCGAGCTGAAGAAGGTCTCTCCAGCAGATGAAGAGAGATGATC 301

QY 290 SerLeuLeuSerProGlnLysLysProArgGluArgAlaGluAspGlyThrGlyThr 309
 |||||
 Db 302 TCTCTCTCTGTCCTCTCAGAGAAGAGCCCGAGGAGGAGAGAGAGAGAGAGAGAGAG 361

QY 310 ValAlaIleSerAspIleGluAspAspSerGlyGluLeuSerArgAspSerValTrpGly 329
 |||||
 Db 362 GTTGCTATCTCCGATATAGAAGATGACTCTCGGGAACCTGACAGAGACAGCGTGTGGGC 421

QY 330 LeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIle 349
 |||||
 Db 422 CTTTCTGTGACACTGTGAGAGAGAGCGCTGACAAACAGCATCAGGAAACAGTGGGAAT 481

QY 350 LeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGluGlyLeu 369
 |||||
 Db 482 TTGAAAAGTCATGTAGAAAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGCGCT 541

QY 370 AsnGluGluAspValIleSerArgGlnAspHisGluGlnGluThrGluLysLeuGluLeu 389
 |||||
 Db 542 AATGAGGAGGAGCGTCATCTCAGCAACAGACCATGAGCAAGAGACTGAGAAACTGGAGCTG 601

QY 390 GluIleGluArgCysLysGluMetIleLysAlaGlnGlnGlnLeuLysGlnGlnLeu 409
 |||||
 Db 602 GAGATTGAGCGGTGTAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTTACACAGCAGCTG 661

QY 410 AlaThrThrCysAspAspThrThrSerLeuLeuArgAspCysTyrTyrLeuLeuGluGlu 429
 |||||
 Db 662 GCCACCGCTGTGATGATGACACCCTCCTCTGTCGAGAGACTGTTACTTGTCTGGAAGAA 721

QY 430 LysGluArgLeuLysGluGluTrpThrLeuPheLysGluGlnLysLysAsnPheGluArg 449
 |||||
 Db 722 AAGGAACGCTTAAAGAGAGTGGACCCCTTTTAAAGAGCAAAAAAGAAATTTTGAGAGA 781

QY 450 GluArgArgSerPheThrGlu 456
 |||||
 Db 782 GAAAGGGGAAGCTTTACAGAA 802

RESULT 14
 CX562093 834 bp mRNA linear EST 12-JAN-2005
 LOCUS UI-M-FC0-ctz-g-10-0-UI.r1 NIH_BMAP_F00 Mus musculus cDNA clone
 DEFINITION IMAGE:30699705 5', mRNA sequence.

ACCESSION CX562093

VERSION CX562093.1 GI:57589122

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 834)

REFERENCE NIH-MGC <http://mgs.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

FEATURES

Location/Qualifiers

source

1. .834

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:30699705"

/tissue_type="whole brain"

/dev_stage="embryo 12.5 dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH BMAP FCO"

/notes="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction.
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA
tail, is TGAGAGAGCC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemm Chin, Ph.D., program coordinator."

ORIGIN

Alignment Scores:

Pred. No.: 7,076-125 Length: 834
Score: 1356.00 Matches: 271
Percent Similarity: 98.2% Conservative: 1
Best Local Similarity: 97.8% Mismatches: 5
Query Match: 42.8% Indels: 1
DB: 9 Gaps: 0

US-10-644-084-2 (1-615) x CX562093 (1-834)

QY 91 LysArgGluLeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuValLeuGln 110
||| :|||
Db 4 AAGAGAGAATAAATATAGTCGCTGTTCTGAACGTATGAACGAGCTGCTGCTTCAG 63

QY 111 ArgLysAsnLeuAlaGlnGluSerValGluThrGlnAsnLeuLysLeuGlySerArg 130
||| :|||
Db 64 CGGAAGAAGCTGCTGCCAGGAGAGCGTGGAGACACAGAACTTGAAGCTGGGCAATGAC 123

QY 131 MetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGluThrSerArgArg 150
||| :|||
Db 124 ATGGACCACCTGCAGAGCTGTACTAGCCAAACTTAAAGAGCAGTTGGAAACGTCCAGCCGG 183

QY 151 GluMetIleGLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHis 170
||| :|||
Db 184 GAGATGATCGGGCTTCAAGAGAGAGACAGCAGCTCAGTGCAGAAACAGGAGTTTGCAT 243

QY 171 GlnLeuLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleIleAlaSerArg 190
||| :|||
Db 244 CAGCTCCTGAAGAATCAGAAAGATGAGGTACAAAAATTACAAATATATACAGCCAGCCGG 303

QY 191 AlaThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGlu 210
||| :|||
Db 304 GCTACTCAGTATATCATGATGTGAAGAGAGAGGAGCGGTGAATATATTAAGCTAAGGAG 363

QY 211 ArgLeuHisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsn 230
||| :|||
Db 364 CGCTGTCATCAGCTCGTTATGAACAGAGAGGATAAAAAATAGCCATGGATGTTTAAAT 423

QY 231 TyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArg 250
||| :|||
Db 424 TATGTGGTCTGAGCTGATGGCAACGAGGCTCATGGAGGACTGACAAAACAGAAAGCCAGG 483

QY 251 AsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIle 270
||| :|||
Db 484 AATGAAGATGAGATGTACAAATTCCTGTTGAATGATTATGATACCCGAGAGCAGATC 543

QY 271 LeuMetGluAsnAlaGluLeuLysValLeuGlnGlnMetLysLysGluMetIleSer 290
||| :|||
Db 544 CTGATGGAGAACGCGAGCTGAAGAAGGCTCCTCCAGCAGATGAAGAAGGAGATGATCTCT 603

QY 291 LeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGlyThrGlyThrVal 310
||| :|||
Db 604 CTCCTGTCTCTCAGAAAGAGAA-CCAGGGAAGAGCAGAGGACGCGCAGGACTGTT 662

QY 311 AlaIleSerAspIleGluAspSerGlyGluLeuSerArgAspSerValTrpGlyLeu 330
||| :|||
Db 663 GCTATCTCCGATATAGACATGACTCTGGGAACTGAGCAGACAGACGCTGTGGGCTT 722

QY 331 SerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeu 350
||| :|||
Db 723 TCTCTGTACACTGTGAGAGAGCAGCTGACANACAGCATCAGGAAACAGTGGGAATTTTG 782

QY 351 LysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGlu 367
||| :|||
Db 783 ANAAGTCATGTAGAAAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAG 833

RESULT 15
CF949831
LOCUS
DEFINITION
IMAGE:30632061 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 840)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. .840
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30632061"
/tissue_type="Upper Head"
/dev_stage="9.5 and 10.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP HJ0"
/note="Organ: Head; Vector: pYX-Asc; Site_1: EcoR I;

Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): "Gene Discovery in the Developing Mouse Nervous System", supported by National Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:

Pred. No.:	3,216-124	Length:	840
Score:	1349.50	Matches:	263
Percent Similarity:	98.9%	Conservative:	0
Best Local Similarity:	98.9%	Mismatches:	2
Query Match:	42.6%	Indels:	1
DB:	5	Gaps:	1

US-10-644-084-2 (1-615) x CF949831 (1-840)

Qy	350	LeuLysSerHisValGluLysLeuAspAenGlnAlaSerLysValHisSerGluGlyLeu	369
Db	1	TTGAAAGTCATGTAGAAAACCTCGATAACCAAGCTTCGAGGTACACTCAGAGGGCCTT	60
Qy	370	AenGluGluAspValIleSerArgGlnAspHisGluGlnGluThrGluLysLeuGluLeu	389
Db	61	AAATGAGGAGGACGTCATCTCAGCAACAAGACCATGAGCAAGAGACTGAGAAACTGGAGCTG	120
Qy	390	GluIleGluArgCysIleGluMetIleLysAlaGlnGlnGlnLeuGlnGlnGlnLeu	409
Db	121	GAGATTGAGCGGTGTAAGAGATGATCAAGGCTCAGCAGCAGCTCTTA---CAGCAGCTG	177
Qy	410	AlaThrThrCysAspAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGlu	429
Db	178	GCCACCACGTGTGATGATGACACCACTCCTGTTGCGAGACTGTTACTTGTCTGGAGAA	237
Qy	430	LysGluArgLeuLysGluGluTrpThrLeuPheLysGluGlnLysAsnPheGluArg	449
Db	238	AAAGGAACGCTTTAAAGAGAGTGGACCCCTTTTAAAGAGCAAAAAAGAAATTTTGAGAGA	297
Qy	450	GluArgArgSerPheThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGlu	469
Db	298	GAAGGCGGAAGCTTTACAGAAGCTGCCATTCGATTGGGGTTGGAGAGAAAGCGTTTGAA	357
Qy	470	GluGluArgAlaSerTrpValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGln	489
Db	358	GAAGAGCGAGCCAGCTGGGTAAAGCAGCAGTTTTAAACATGACGAACTTTGACCACCAG	417
Qy	490	AenSerGluAsnValLysLeuPheSerAlaPheSerGlySerSerAspProAspAsnLeu	509
Db	418	AACTCAGAAAATGTGAAACTTTTCAGTGCCTTCTCAGGAAGTCTCTGATCCAGACAACTT	477
Qy	510	IleValHisSerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAla	529
Db	478	ATAGTCCACTCACGGCCACGGCAAAAGAGCTACACAGTGTGGCTTAATGGGGTGCAGCT	537
Qy	530	CysThrSerLysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArgGln	549
Db	538	TGCACATCAAAACTGACTAAATCTCTTCTGCCTCACCTTCTACTTCAGACTTTCGCCAG	597
Qy	550	ThrHisSerCysValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGlu	569
Db	598	ACACATTTCATGTGTCTGAAACACACAGTTCACACAGTGTGCTGAAATATAACTCTCGAAGAA	657
Qy	570	SerLysProSerGluValAlaArgGluSerThrAspGlnLysTrpSerValGlnSerArg	589
Db	658	AGTAAACCAAGTGAGGTTTGCAGAGAAAGCAGCGGATCAGAAAGTGAGCGCTGAGTCGAGG	717

Qy	590	ProSerSerArgGluGlyCysTyrSerGlyCysSerSerAlaPheArgSerAlaHisGly	609
Db	718	CCCAGCTCGCGGAGGGGTGCTACAGCGGATGCTCTCGGNCCTTCANGAGCGCTCACCGN	777
Qy	610	AspArgAspAspLeuPro	615
Db	778	GACCGAGATGACTTACCT	795

Search completed: June 14, 2006, 01:45:29
Job time : 7935 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2006, 22:24:24 ; Search time 498 Seconds
(without alignments)
10114.508 Million cell updates/sec

Title: US-10-644-084-1
Perfect score: 2692
Sequence: 1 cgtaggagagcaggagc.....ttgattatggagtaatg999 2692

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 93554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /EMC Celerra_IDS3/ptodata/2/ina/1 COMB.seq:*
- 2: /EMC Celerra_IDS3/ptodata/2/ina/5 COMB.seq:*
- 3: /EMC Celerra_IDS3/ptodata/2/ina/6A COMB.seq:*
- 4: /EMC Celerra_IDS3/ptodata/2/ina/6B COMB.seq:*
- 5: /EMC Celerra_IDS3/ptodata/2/ina/7 COMB.seq:*
- 6: /EMC Celerra_IDS3/ptodata/2/ina/H COMB.seq:*
- 7: /EMC Celerra_IDS3/ptodata/2/ina/PCTUS COMB.seq:*
- 8: /EMC Celerra_IDS3/ptodata/2/ina/PP COMB.seq:*
- 9: /EMC Celerra_IDS3/ptodata/2/ina/RE COMB.seq:*
- 10: /EMC Celerra_IDS3/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1258.8	46.8	2272	3	US-10-104-047-1497 Sequence 1497, Ap
2	297.2	11.0	428	3	US-09-513-999C-3004 Sequence 3004, Ap
3	73.6	2.7	7218	2	US-08-232-463-14 Sequence 14, Appl
4	58.2	2.2	198	3	US-09-513-999C-29101 Sequence 29101, A
5	55.8	2.1	2093	3	US-10-104-047-1566 Sequence 1666, Ap
6	50.8	1.9	1141	3	US-09-806-708B-22 Sequence 22, Appl
7	48.8	1.8	5883	3	US-09-949-016-5001 Sequence 5001, Ap
8	48.8	1.8	28806	3	US-09-949-016-13217 Sequence 13217, A
9	46.6	1.7	1312	3	US-09-976-594-886 Sequence 886, App
10	46	1.7	3489	2	US-08-728-323A-1 Sequence 1, Appl
11	46	1.7	3489	3	US-09-298-568-1 Sequence 1, Appl
12	46	1.7	3489	3	US-09-410-399-1 Sequence 1, Appl
13	46	1.7	3489	3	US-09-894-273-1 Sequence 1, Appl
14	46	1.7	32207	2	US-08-770-379-20 Sequence 20, Appl
15	46	1.7	32207	3	US-08-757-669A-20 Sequence 20, Appl
16	46	1.7	32207	3	US-09-230-371A-20 Sequence 20, Appl
17	44.4	1.6	543	10	5273901-6 Patent No. 5273901
18	44.4	1.6	832	3	US-09-621-976-2813 Sequence 2813, Ap
19	43.6	1.6	16442	3	US-08-781-891-208 Sequence 208, App
20	43.6	1.6	16442	3	US-09-618-166-208 Sequence 208, App
21	43.4	1.6	2357	3	US-10-104-047-1780 Sequence 1780, Ap
22	43.4	1.6	15252	3	US-09-949-016-13584 Sequence 13584, A
23	43	1.6	3337	2	US-08-072-610-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-10-104-047-1497
; Sequence 1497, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cdna
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1497
; LENGTH: 2272
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1497

Query Match.	46.8%	Score 1258.8;	DB 3;	Length 2272;
Best Local Similarity	77.2%;	Pred. No. 0;		
Matches 1651;	Conservative 0;	Mismatches 337;	Indels 150;	Gaps 4
Qy	1	CGTAGGAGAGTGACAGGAGCTGTTGTAAGCGTCGACGACCTGAGCGGCTCTCTCAGGTAT	60	
Db	145	CATAGAGAGTGACAGGAGCTAGACTAAATGTTTAACTGCTGAACTAGTCTCTCAGGTAT	204	
Qy	61	CCTGGCTCTGGAACCTTCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTGAC	120	
Db	205	CCTGGCTCTAGAGATTGCTATGGGAGATTGGATGACTGTTACAGATC-----C	252	
Qy	121	AGAAAAACAAAAATCTCTCTCAATATACCTCAGAAAAACAAAGATGCTCCGTCACGTTTGTGA	180	
Db	253	AGAAAGCAAACTATCTCTCAATATACCTCAGAAAAACAAAGATGCTCCATCAAGTTTATA	312	
Qy	181	CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTATCCAAAAACGTCATGGTGTGTTTCGG	240	
Db	313	CTCACAGCAAGTGTATGTTCTTCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTTCAG	372	
Qy	241	TGCTCTTCTGCACAGGAGAGAACTTGAACAAAGTATTTCTCTATCTTGTATCAGGAGCTGAC	300	
Db	373	TGCCTTCTGCACAGAAAGATAATATTGAACAGAGTATCTCATATCTTGTATCAGGAATTTGAC	432	
Qy	301	CACCTTCGGGTTTTCTTCTTGTATGAAGAAATCCAAAGATTAAGAGCGCAAGAGAGAATTT	360	

Db 433 TACTTTTGGTTTTCTTCATTATATGAAGATCCAAAGGTAAAGAGACAAAGAGAGAGTT 492
Qy 361 AAATATAGTCGCTGTTCTGAACCTGTATGAACGAGCTGCTGCTGCTTTCAGCGGAAGAACCT 420
Db 493 AAATATAGTAGCTGTACTAAATTGATGAATGAGCTGCTTGTGCTTCAGCGGAAGAACCT 552
Qy 421 GCTGCCCAGGAGAGCGTGGAGACACAGAACTTCAAGCTGGGACGTGACATGGACACCT 480
Db 553 TCTAGCTCAGGAAATGTGGAGACACAGAAATTTGAAGCCGGGAAGTATATGGACCATCT 612
Qy 481 GCAGAGCTGTACGCCAAACTTAAGGAGCAGTTGGAAACGTCCTCAGCGGGAGATGATCGG 540
Db 613 ACAGAGCTGTACTCAAACTTAAGAACACTCGGAACCTCCAGGAGGGAATGATGG 672
Qy 541 GCTTCAAGAGAGACAGGCGAGCTGCGAGTGAAGAACAGGAGTTTGCATCAGCTCCTGAA 600
Db 673 GCTTCAGGAAGAGACAGACAGTACAAATGATGAAGAACAGGAATTTGCATCAGCTACTAAA 732
Qy 601 GAATGAGAAAGATCAGGTACAAAATTTACAAAATATCATAGCCAGCGGGCTACTCAGTA 660
Db 733 GAATGAGAAAGATCAGGTGCAAAAATTTACAAAATATCATTCGAAGTCAAGCTACTCAGTA 792
Qy 661 TAATCATGATGTGAAGAGGAGGAGCGGTGAATATAAAGCTAAAGGAGCGCCTGCAATCA 720
Db 793 TAATCATGATGTGAAGAGGAGGAGCGGTGAATATAAAGCTAAAGGAGCGCTACTACATCA 852
Qy 721 GCTGTTATGAACAAGAGGATAAACATAGCCATGATGTTTTTAATATGTTGGGTG 780
Db 853 ACTTGTATGAACAAGAGGATTAAGAAATAGCTATGACATTTTGAATATGTTGGGAG 912
Qy 781 AGCTGATGGCAACCGAGGCTCATGGAGCTGACAAACAGAACAGAGGATGAAGATGA 840
Db 913 AGCTGATGGAAAGAGAGGCTCCTGGAGACTGGTAAACTGAGCCAGGAATGAAGATGA 972
Qy 841 GATGTACAAAATCTGTTGAATGATTTAGTACCCGAGAGAGAGATCTCTGATGGAGAA 900
Db 973 AATGTATAAATCTTTGAATGATTAATATCGTCAGAAACAAATCTTAATGGGAAA 1032
Qy 901 CGCGAGCTGAAGAGGCTCTCAGCAGATGAAGAGAGATGATCTCTCTCTGCTCC 960
Db 1033 TGCAGAACTTAAGAGGTTCTTCAACAAATGAAAGAGGAATGATTTCTCTCTCTCTCC 1092
Qy 961 TCAGAAAGAGAGCCAGGGAAAGAGCAGAGAGCGGACAGCGCACTGTTGTATCTCCGA 1020
Db 1093 CCAAGAGAGAAACCTAGAGAAAGATGATGATAGTACAGAACTGT---TATTTCCGA 1149
Qy 1021 TATAGAAGATGACTCTGGGAACTGAGCAGAGACAGCGTGTGGGCGCTTCTCTGTGACAC 1080
Db 1150 TGTGAAGAGATGCGCGGGAACTAAGCAGAGAGATGATGTGGGACCTTTCTCTGTGAAC 1209
Qy 1081 TGTGAGAGAGAGCTGACAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAGTCAATGT 1140
Db 1210 TGTGAGAGAGAGCTTACAAACAGCATCAGAAACAGTGGAGAAATTTTGAAGATCAATGT 1269
Qy 1141 AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGGCTTAAATGAGGAGACGT 1200
Db 1270 AGAAAGCTGTATAACCAAGTTTCAAGGTACACTGGAAGTTTAAATGATGAAGATGT 1329
Qy 1201 CATCTCAGCAAGAACCATAGCAAGAGACTGAGAAACTGGAAGTGGAGATTTGAGCGGTG 1260
Db 1330 AATCTCAGCAAGAACCATGAACAGAACTGAAAACTCGAGTTAGAAATTCAGCAGTG 1389
Qy 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTTTACAGCAGCAGCTGGCCACCAGTGTGA 1320
Db 1390 TAAAGAAATGATTAACCTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCAATGA 1449
Qy 1321 TGATGACACCACTCAGCTGTTGCGAGACTGTTACTTGTCTGGAAGAAAGAACGCTTAA 1380
Db 1450 TGATGATACCACTTCACTATTACGAGACTGTTATTTGTTGGAAGAAAGGACGCTCTCA 1509
Qy 1381 AGAAGAGTGGACCTTTTAAAGAGCAAAAAAGAAATTTTGAAGAGAGAAAGCGCAAGCTT 1440
Db 1510 AGAAGATGGTCCCTTTTAAAGAGCAGAAAAAGAAATTTTGAAGAGGAGAGACGAAGCTT 1569

Qy 1441 TACAGAGCTGCCATTGATTGGGTTGGAG----- 1471
Db 1570 TACAGAGCCGCTATTGCGCTGGGATTTGGAGATTTGGCTTTCTCTGCAAGTGTTTTACA 1629
Qy 1472 ----- 1471
Db 1630 CAAGAAATCTCTTTTGGGTATCCATTCATCCACGTCAGTAGGTGTGGTGTGAGAAG 1689
Qy 1472 -----AGAAAGGCGTTTGAAGAGAGGAG 1496
Db 1690 ACAGTGAACCAAGACTATATTGCTGTTTTAAAAAGAAAGGCAATTTTGAAGAAAGAGAG 1749
Qy 1497 CCAGCTGGGTAAAGCAGCAGTTTTTAAACATGACGAACTTTTGACCAACAGAACTCAGAAA 1556
Db 1750 CCAGTTGGTTAAAGCAGCAGTTTCTAAATATGACTACCTTTGACCACAGAACTCAGAAA 1809
Qy 1557 ATGTGAAACTTTTTCAGTGCCTTCTCAGGAAATGTTCTGATCCAGACAATCTTTATAGTCCACT 1616
Db 1810 ATGTGAAACTTTTTCAGTGCCTTCTCAGGAAATGTTCTGATTTGGGACAACTTTATAGTGCAC 1869
Qy 1617 CAGGCCACGGCAAAAGAGCTACAGAGTGTGGCTAATGGGTGCCAGCTTGCACATCAA 1676
Db 1870 CGAGGAGCGCGCAAAAGAGCCTCACAGTGTGTCTAATGGGTCTCCAGTTTGCATGTCTA 1929
Qy 1677 AACTGACTAAATCTCTTCTGCGCTCACCTTCTACCTTCAGACTTTTCGGCAGACACATTCAT 1736
Db 1930 AACTTACTAAATCTCTTCTGCTTCCCTTCCACTTCCACTTCAGACTTTTGCAGACAGCTTCT 1989
Qy 1737 GTGTGTCTGAACAGTTCATCAGTGTGTGTAATATACTCTCTGAAGAAAGTAAACCAA 1796
Db 1990 GCATATCTGAACATAGTTCAATCAATGTAATAAATACTGCTGAAGAAATTAACCAA 2049
Qy 1797 GTAGGTTTGAAGAAAGCAGCATCAGAGTGTGGAGCTGAGTCCAGGCGCCAGCTCGC 1856
Db 2050 ATCAGGTTGGAGGAGAACGTACAAATCAAAATTTGGAGTGTGGGTCAAGACCTGGATCAC 2109
Qy 1857 GGAGGGGTCTACAGCGGATGCTCTCGGCTTCAGGAGCGCTCACGGGACCGAGATG 1916
Db 2110 AGAAGGTTGCTATAGTGGATGCTCTTCTGAGCTACACAAATTTCTCATGTAGAAAAGATG 2169
Qy 1917 ACTTACCTTAAATGTGCGGCTGCAGTGTGTTTCCCAGATGTGCGGTAGAGG----- 1968
Db 2170 ACTTACCTTAGACATGTGGACTGGAATTTTTTTTCAATTAATGTGTTTCATCAAGTTTCACAT 2229
Qy 1969 ----AGTTGACAGGTTGTAGCATAAAGTCAGTCGCT 2003
Db 2230 CTAAGTTGAACAGGGGTGTGTCTATAAAGTCAGTTATCT 2267

RESULT 2

US-09-513-999C-3004
; Sequence 3004, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 3004
; LENGTH: 428
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS

GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5001
; LENGTH: 5883
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5001

Query Match 1.8%; Score 48.8; DB 3; Length 5883;
Best Local Similarity 48.9%; Pred. No. 0.011;
Matches 131; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 342 AAGCGCAAGAGAGAGATTAATATAGTCGCTGTTCTGNACTGTATGAACGAGCTGCTCG 401
DB 4094 AGGTGGCCGACATGAATAAAGAGATGGAGACAGTGTGGGGTGCCTGGAAACTGCTGAGG 4153

QY 402 TCGTTTCAGCGGAAGAACTGCTGGCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGG 461
DB 4154 AGGTGAAGAGGAGCTCCAGAGGACCTGGAGGGCTGAGCCAGCGGCACAGGAGGAAGG 4213

QY 462 GCAGTGACATGACCACTGCTGAGAGCTGCTACGCCAAACTTAAGAGAGCAGTTGGAAACGT 521
DB 4214 TGGCGCGCTACGACAGCTGGAGAGACCAAGACGCGGTGAGCAGGAGCTGGACGACC 4273

QY 522 CCAGCGGGAGATGATCGGGCTTCAAGAGAGACAGGAGCTGAGTGCAGTGCAGAACAGGA 581
DB 4274 TGCTGGTGACCTGGACCAACCGCCAGAGCGCGTGCACCTGGAGAGAACGACGAGA 4333

QY 582 GTTTCATCAGCTCTCTGAAGAAATGAGAA 609
DB 4334 AGTTTGACCACTCTCTGGCGGAGGAGAA 4361

RESULT 8
US-09-949-016-13217/c
; Sequence 13217, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13217
; LENGTH: 28806
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13217

Query Match 1.8%; Score 48.8; DB 3; Length 28806;
Best Local Similarity 47.4%; Pred. No. 0.03;

Matches 146; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 881 AAGCAGATCTCGATCGAGAACCGGAGCTGAAGAAAGTCTCTCCAGCAGATGAAGAAGGAG 940
DB 25834 AAGGAG 25775

QY 941 ATGATCT 1000
DB 25774 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 25715

QY 1001 GGCACCTGTTGCTATCTCCGATATAGAAGTACTCTCTGGGAACTCTGAGCAGACAGCCGTG 1060
DB 25714 CAGACAGAGAGGAG 25655

QY 1061 TGGGGCCTTTCTCTGTACACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1120
DB 25654 AGAGAGAGACAGAGGGTGCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 25595

QY 1121 AGAATTTTGAAGAGTCACTGTAGAAAAAACTCGATACCAAGCTTCGAAGGTACACTCAGAG 1180
DB 25594 AGGATACATGAATAATGATAGCAGATGAAATTCATACCCAGAGACCCAGGAGACACAGAG 25535

QY 1181 GGCCTTAA 1188
DB 25534 AAACAGAA 25527

RESULT 9
US-09-976-594-886
; Sequence 886, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 886
; LENGTH: 1312
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 981037.1
US-09-976-594-886

Query Match 1.7%; Score 46.6; DB 3; Length 1312;
Best Local Similarity 52.9%; Pred. No. 0.017;
Matches 100; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 376 TCTGAACGTATGAACCGAGCTGCTGCTCTTTCAGCGGAAGAACTGTGTCGCCAGGAGAG 435
DB 640 TCACCTACTTTCTGCCCAACCTCGACCTCTTTTCAGGGCAAGGCCCATTCGGCCCTGGAGAG 699

QY 436 CGTGAGACACAGNACTTGAAGCTGGGAGTGCATGAGACCATCTGCAGAGCTGCTAGCC 495
DB 700 CGTGCCCAAGCAGACCTGGAGGTTGGCCAGGAAATTTCTCACTCAATCCCAAAAGCCTTGA 759

QY 496 CAAACTTAAGGAGCAGTTGGAAACGTCAGCGGGGAGATGATCGGGCTTCAAGAGAGAGA 555
DB 760 CAACTATAGGGTCTGGGACTCTCTTTGAAAGGACACAAACGGGGTGTCTCTCTAGA 819

QY 556 CAGCAGCT 564
DB 820 CACAACT 828

RESULT 10

US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
US-08-728-323A-1
Query Match 1.7%; Score 46; DB 2; Length 3489;
Best Local Similarity 42.7%; Pred. No. 0.048;
Matches 235; Conservative 0; Mismatches 315; Indels 0; Gaps 0;
Qy 393 AGCTGCTCGTCTTACGGGAAGAACCTGCTGGCCCCAGGAGCGGTGGAGACAGAACT 452
Db 1904 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC 1963
Qy 453 TGAAGCTGGCAGTGACATGCACCTGCAGAGCTGCTAGCCAACTTAAGGAGCAGT 512
Db 1964 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC 2023
Qy 513 TGAAGAACTCCAGCGGGGAGATGATCGGGCTTCAAGAGAGAGACAGGCGCTGCAGTGCA 572
Db 2024 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC 2083
Qy 573 AGAACAGAGTTTGCATCAGCTCTGAAGAATGAGAAAGATGAGGTACAAAAATTACAA 632
Db 2084 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC 2143
Qy 633 ATATCATAGCCCGGCTACTCAGTATATCATGATGATGATGATGATGATGATGATGATGATGAT 692
Db 2144 AGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC 2203
Qy 693 ATAATAAGCTAAAGGAGCGCTTGCATCAGCTCGTTATGAACAAGAGGATAAAAACATAG 752
Db 2204 AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC 2263
Qy 753 CCATGGATGTTTTTAATTATGTGGTGCAGCTGATGCAACAGGAGGCTCATGGAGGACTG 812

US-09-298-568-1
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestar, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: RHADINO VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1
Query Match 1.7%; Score 46; DB 3; Length 3489;
Best Local Similarity 42.7%; Pred. No. 0.048;
Matches 235; Conservative 0; Mismatches 315; Indels 0; Gaps 0;
Qy 393 AGCTGCTCGTCTTACGGGAAGAACCTGCTGGCCCCAGGAGCGGTGGAGACAGAACT 452
Db 1904 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC 1963
Qy 453 TGAAGCTGGCAGTGACATGCACCTGCAGAGCTGCTAGCCAACTTAAGGAGCAGT 512
Db 1964 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC 2023
Qy 513 TGAAGAACTCCAGCGGGGAGATGATCGGGCTTCAAGAGAGAGACAGGCGCTGCAGTGCA 572
Db 2024 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC 2083
Qy 573 AGAACAGAGTTTGCATCAGCTCTGAAGAATGAGAAAGATGAGGTACAAAAATTACAA 632
Db 2084 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC 2143
Qy 633 ATATCATAGCCCGGCTACTCAGTATATCATGATGATGATGATGATGATGATGATGATGATGAT 692
Db 2144 AGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC 2203
Qy 693 ATAATAAGCTAAAGGAGCGCTTGCATCAGCTCGTTATGAACAAGAGGATAAAAACATAG 752
Db 2204 AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC 2263
Qy 753 CCATGGATGTTTTTAATTATGTGGTGCAGCTGATGCAACAGGAGGCTCATGGAGGACTG 812

Db	2254	AGCAGGACGAGGACGAGGACGAGGAGGACGAGGAGGATTTAGAGGACGAGGACGAGG	2323
QY	813	ACAAACAGAACCCAGGAAATGAAGATGATGATGTTCAAAATTTCTGTGAATGATTATGAGT	872
Db	2324	AGTTAGAGGATCAGGACGAGGATTTAGAGGACGAGGACGAGGATTTAGAGGACGAGGACG	2383
QY	873	ACCGCCAGAACGAGATCCTGATGAGAGAACGCCGAGCTTGAAGAAAGGTCTCTCCAGCAGATGA	932
Db	2384	AGGAGTTAGAGGACGAGGACGAGGACGAGGATTTAGAGGAGGACGAGGACGAGGATTTAGAGGACGAGG	2443
QY	933	AGAAGGAGAT	942
Db	2444	AGCAGGAGTT	2453

```

RESULT 12
US-09-410-399-1
; Sequence 1, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; TITLE OF INVENTION: to Genomic Host DNA
; FILE REFERENCE: UN-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-1

```

[illegible]

```

Db      2384  AGGAGTTAGGAGCAGGAGCAGGAGTGTAGAGGAGCAGGAGTGTAGAGGAGCAGG  2443
QY      933   AGAAGAGAGAT  942
        |||||
Db      2444  AGCAGGAGATT  2453

RESULT 13
US-09-894-273-1
; Sequence 1, Application US/09894273
; Patent No. 6756203
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/894,273
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/109,422
; PRIOR FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-894-273-1

```

Query Match	1.7%	Score 46	DB 3	Length 3489
Best Local Similarity	42.7%	Pred. No. 0.048		
Matches 235	Conservative 0	Mismatches 315	Indels 0	Gaps 0
Qy	393	AGCTGCTCGTCTTTCAGCGGAAGAACTGCTGTCGCCCAGGAGAGCGGTGGAGACACAGAACT	452	
Db	1904	AGCAGCAGGATGACAGCAGCACAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC	1963	
Qy	453	TGAAGCTGGGCAGTGACATCGACCACTGTCAGAGCTGTCTACGCCAAACTTTAAGGAGCAGT	512	
Db	1964	AGCAGCAGGATGACAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC	2023	
Qy	513	TGGAAACGTCCAGGCGGAGATGATCGGCGTTCAGAGAGAGACAGCGCAGCTCGACGTGCA	572	
Db	2024	AGCAGCAGGATGACAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGC	2083	
Qy	573	AGAACCGAGCTTTGCATCAGCTCCTCAAGAAATCAGAAAGATGAGGTACAAAAATTACAAA	632	
Db	2084	AGCAGCAGGATGACAGGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC	2143	
Qy	633	ATATCATAGCCAGCGGGCTTCTCAGTATATCATGATGTGAAGAGAGAGGACGCTGAAT	692	
Db	2144	AGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGC	2203	
Qy	693	ATAATAGCTAAAGGAGCGCTCCATCAGCTCTTATGACACAGAGAGGTAAAAACATAG	752	
Db	2204	AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC	2263	
Qy	753	CCATGGATGTTTTAAATTTATGTGGGTCGACTGATGGCAAAACGAGGCTCATGGAGACTG	812	
Db	2264	AGCAGGAGGAGCAGGAGCAGCAGGAGGAGCAGCAGGAGGATGAGGAGCAGCAGGAGCAGG	2323	
Qy	813	ACAAAACAGAACCCAGGATGGAAGATGAGATGTACAAAATTTCTGTTTGAATGATATGAGT	872	
Db	2324	AGTTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGC	2383	
Qy	873	ACCGCCAGAAGCAGATCCCTGATGGAGAAACGCGAGCTGAAGAAGGCTCTCCAGCAGATGA	932	
Db	2384	AGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGG	2443	
Qy	933	AGAAGGAGAT	942	
Db	2444	AGCAGGAGTT	2453	

RESULT 14
US-08-770-379-20/c
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770.379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-770-379-20
Query Match 1.7%; Score 46; DB 2; Length 32207;
Best Local Similarity 42.7%; Pred. No. 0.21;
Matches 235; Conservative 0; Mismatches 315; Indels 0; Gaps 0;
QY 393 AGCTGCTCGTCTCAGCGGAAGAACCTGCTGGCCCGAGGAGCGTGGAGACACAGAACT 452
DB 20093 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC 20034
QY 453 TGAAGCTGGGCGAGTGCATGCGGGCTTCAAGAGAGAGACAGCGGCGTGCAGTGCA 512
DB 20033 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC 19974
QY 513 TGGAAACCTCCAGGCGGAGATGATCGGGCTTCAAGAGAGAGACAGCGGCGTGCAGTGCA 572
DB 19973 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC 19914
QY 573 AGACAGGAGTTTGCATCAGCTCTCGAAGATGAGAAAGATGAGGTACAAAATTACAAA 632
DB 19913 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC 19854
QY 633 ATATCATAGCCAGCGGGCTACTCAGTATATCATGTGTAAGAGGAGGAGCGTGAAT 692
DB 19853 AGGATGACAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGC 19794
QY 693 ATAATAAGCTAAAGGAGCGCTTGCATCAGCTCGTTATGAAACAAGAGGATAAAAACATAG 752

DB 19793 AGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGGAGC 19734
QY 753 CCATGATGTTTTAAATTATGTGGTTCGAGCTGATGCCAACGAGGCTCATGGAGGACTG 812
DB 19733 AGCAGGAGGAGCAGGAGCAGCAGGAGGAGCAGGAGGAGGAGGAGGATTTAGAGGAGCAGGAGCAGG 19674
QY 813 ACAAACAGAAAGCCAGGAATGAAGATGAGATGTACAAAATTCCTGTTGAATGATTATGAGT 872
DB 19673 AGTTAGAGGATCAGGAGCAGGAGTTTAGAGGAGCAGGAGGAGGATTTAGAGGAGCAGGAGC 19614
QY 873 ACCGCCAGAAAGCAGATCCTGATGGAGAAACCGGAGCTGGAAGAGGTCTCTCCAGCAGATGA 932
DB 19613 AGGAGTTAGAGGAGCAGGAGCAGGAGTTTAGAGGAGCAGGAGGAGGAGGATTTAGAGGAGCAGG 19554
QY 933 AGAAGAGAT 942
DB 19553 AGCAGGAGTT 19544
RESULT 15
US-08-757-669A-20/c
; Sequence 20, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757.669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20
Query Match 1.7%; Score 46; DB 3; Length 32207;
Best Local Similarity 42.7%; Pred. No. 0.21;
Matches 235; Conservative 0; Mismatches 315; Indels 0; Gaps 0;
QY 393 AGCTGCTCGTCTCAGCGGAAGAACCTGCTGGCCCGAGGAGCGTGGAGACACAGAACT 452
DB 20093 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC 20034
QY 453 TGAAGCTGGGCGAGTGCATGCGACCCTGCTGAGAGCTGTACGCCAACTTAAGAGCAGT 512

[illegible]

Search completed: June 13, 2006, 23:33:10
Job time : 501 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 12, 2006, 19:14:34 ; Search time 51 Seconds
(without alignments)
1055.516 Million cell updates/sec

Title: US-10-644-084-2

Perfect score: 3165

Sequence: 1 MGDWMTVDPVLTENKNLS.....CYSGSSAFSAHGDRDLP 615

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /EMC_Celerra_SIDS3/ptodata/2/iaa/5_COMB.pep:*
- 2: /EMC_Celerra_SIDS3/ptodata/2/iaa/6_COMB.pep:*
- 3: /EMC_Celerra_SIDS3/ptodata/2/iaa/7_COMB.pep:*
- 4: /EMC_Celerra_SIDS3/ptodata/2/iaa/H_COMB.pep:*
- 5: /EMC_Celerra_SIDS3/ptodata/2/iaa/pCTUS_COMB.pep:*
- 6: /EMC_Celerra_SIDS3/ptodata/2/iaa/RE_COMB.pep:*
- 7: /EMC_Celerra_SIDS3/ptodata/2/iaa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2152.5	68.0	504	2	US-10-104-047-3467
2	612.5	19.4	137	2	US-09-513-999C-7081
3	223.5	7.1	359	2	US-09-949-016-6507
4	222.5	7.0	905	2	US-09-248-796A-16333
5	222.5	7.0	2733	2	US-09-949-016-11433
6	220.5	7.0	1780	2	US-09-949-016-6899
7	220.5	7.0	1786	2	US-09-949-016-7880
8	215.5	6.8	2125	2	US-09-919-172-29
9	212.5	6.7	2863	2	US-09-538-092-1252
10	211.5	6.7	1388	1	US-08-685-576-1
11	211.5	6.7	2662	2	US-09-595-684B-31
12	210.5	6.7	1354	2	US-08-685-871-2
13	208.5	6.6	1388	1	US-08-685-576-4
14	208.5	6.6	1388	2	US-09-976-594-296
15	208	6.6	2704	2	US-09-538-092-1260
16	206.5	6.5	1676	2	US-09-949-016-7610
17	206.5	6.5	2186	2	US-09-949-016-10828
18	206.5	6.5	2349	2	US-09-538-092-914
19	204	6.4	1427	2	US-09-538-092-1044
20	203	6.4	3210	2	US-09-538-092-1154
21	203	6.4	3248	1	US-08-353-700-1
22	203	6.4	3248	5	PCR-US95-16216-1
23	200.5	6.3	1958	2	US-10-028-946-4
24	200.5	6.3	2053	2	US-09-964-956-11
25	200.5	6.3	2054	2	US-10-028-946-2
26	200.5	6.3	2066	2	US-09-964-956-9

27	200	6.3	2482	1	US-08-328-254-6	Sequence 6, Appli
28	199.5	6.3	1401	2	US-09-750-590A-2	Sequence 2, Appli
29	199.5	6.3	2053	2	US-10-017-216-2	Sequence 2, Appli
30	197.5	6.2	2107	2	US-09-949-016-7646	Sequence 7646, Ap
31	197.5	6.2	2107	2	US-09-949-016-7647	Sequence 7647, Ap
32	197.5	6.2	2115	2	US-09-296-662-33	Sequence 33, Appl
33	195	6.2	860	2	US-10-037-417-59	Sequence 59, Appl
34	195	6.2	1695	2	US-08-866-108A-15753	Sequence 15753, A
35	195	6.2	1972	2	US-08-875-435B-4	Sequence 4, Appli
36	194.5	6.1	1979	2	US-09-949-016-6468	Sequence 6468, Ap
37	194.5	6.1	2047	2	US-09-949-016-7404	Sequence 7404, Ap
38	194	6.1	1024	2	US-09-270-767-44973	Sequence 44973, A
39	193	6.1	994	2	US-09-949-016-6779	Sequence 6779, Ap
40	193	6.1	1055	2	US-09-949-016-9776	Sequence 9776, Ap
41	193	6.1	1234	2	US-09-592-054-8	Sequence 8, Appli
42	193	6.1	3878	2	US-09-914-259-11	Sequence 11, Appl
43	191.5	6.1	1388	2	US-09-572-191-2	Sequence 2, Appli
44	191.5	6.1	1388	2	US-09-723-262-2	Sequence 2, Appli
45	191.5	6.1	1388	2	US-09-723-219-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-104-047-3467
; Sequence 3467, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cdna
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3467
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3467

Query Match	68.0%;	Score 2152.5;	DB 2;	Length 504;
Best Local Similarity	90.3%;	Pred. No. 9.2e-157;		
Matches 419;	Conservative 20;	Mismatches 20;	Indels 5;	Gaps 2;
Qy	1	MGDWMVTDPVLTCTENKNLSQYTSKMSPSLSYQQVLCSSVPLSKNVHGVFGVCTGGE	60	
Db	1	MGDWMVTDP-----ESKTSQYTSKMSPSLSYQQVLCSSVPLSKNVHSPFSAFCTED	56	
Qy	61	NIEQISYLDQELTTFGFPSPSYEESKSKRELNIVAVLNCMELLVLORKNLLAOSV	120	
Db	57	NIEQISYLDQELTTFGFPSPSYEESKSKRELNIVAVLNCMELLVLORKNLLAQN	116	
Qy	121	ETQNLKGSMDHLSQYAKLKEQLETSRREMIGIQERDRQOCNRSIHLQLLKNKEDV	180	
Db	117	ETQNLKPSGMDHLSQYSLKKEQLETSRREMIGIQERDRQOCNRLHQLLKNKEDV	176	
Qy	181	OKLQNIISARATQYNHDKRKEREYVKUKERHLQVLMMKKKNIAMVNLVYGRADGKRG	240	
Db	177	OKLQNIISARATQYNHDKRKEREYVKUKERHLQVLMMKKKNIAMVNLVYGRADGKRG	236	
Qy	241	SWRTGKTARNEDEMYKILLNDYEROKQILMENAELKKVLOOMKEMISLLSPKKKPR	300	
Db	237	SWRTGKTARNEDEMYKILLNDYEROKQILMENAELKKVLOOMKEMISLLSPKKKPR	296	
Qy	301	EBAEDGTGTAISDIEDDSGELSDVSWGLSCDTVREQLTNSIRKQWILKSHVEKLDNQ	360	
Db	297	ERVDDSTGTV-ISDVEEDAGELSRWSMDLSCETVREQLTNSIRKQWILKSHVEKLDNQ	355	
Qy	361	ASKVHSEGLNEEDVISRODHEQTEKLELETERCKEMIKAOQQLLQOQLATTCTDDDTTSL	420	

Db 356 VSKVHLEGNDENVISRDHEQETEKLELEIQCKEMIKTQQLQQLATAYDDTTSL 415
Qy 421 LRDCYLLEEKERLKEWTLFKQCKNFERERSFTAAIRLGL 464
Db 416 LRDCYLLEEKERLKEWTLFKQCKNFERERSFTAAIRLGL 459

RESULT 2
US-09-513-999C-7081
; Sequence 7081, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 7081
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-513-999C-7081

Query Match 19.4%; Score 612.5; DB 2; Length 137;
Best Local Similarity 87.0%; Pred. No. 1.8e-39;
Matches 120; Conservative 10; Mismatches 7; Indels 1; Gaps 1;
Qy 272 MENAELKVKLQCKEMISLLSPQKKKPRERAEDGTGTVASIDIEDDGSGLSRDWSGLS 331
Db 1 MENAELKVKLQCKEMISLLSPQKKKPRERVDSTGTV-ISDVEDAGLSRESMDLS 59
Qy 332 CDTVREQLTNSIRKQWRIILKSHVEKLDNOASKVHSEGLNEEDVISRDHEQETEKLELEI 391
Db 60 CETVREQLTNSIRKQWRIILKSHVEKLDNOASKVHSEGLNEEDVISRDHEQETEKLELEI 119
Qy 392 ERCKEMIKAQQLQQL 409
Db 120 QCKEMIKTQQLQQL 137

RESULT 3
US-09-949-016-6507
; Sequence 6507, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6507
; LENGTH: 3259
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6507

Query Match 7.1%; Score 223.5; DB 2; Length 3259;
Best Local Similarity 19.9%; Pred. No. 1e-07;
Matches 147; Conservative 136; Mismatches 218; Indels 237; Gaps 34;
Qy 15 ENKLSQYTSKMS-----PSSLYSQVLCSSVP-LSKNVHGVGVFCTGENTIQ 64
Db 1225 ENENIGDQLQIQVRESIDGKLPT--DQESCSSTPGLEELFKATEQHHTQPVLES 1282
Qy 65 SI-----SYLDOELTTFGFPSIYE---ESKSKEAKRELNIIVAVLNCMELLVLQKKNLLA 116
Db 1283 NLCPDWPSPHSDASALQGTVAQIKAKIEAEKVLELKVSTTSBELTKKSEVQQL 1342
Qy 117 QESVETQNLKLS--DMDHLQSCYAK-LKEQLETSRRREMIGLQERDROLQCKNRLSHQLL 173
Db 1343 QEINKQGLETESLKTVSHEAEVHAESLQKLESSQLQIAGL-EHLRELQPK-----1393
Qy 174 KNEKDEVOKLQNIITASRATQYNHDVKERYNKLERLHQLVMNKKDKNTAMDVLNVYG 233
Db 1394 ---LDELQKL-----ISKKEEDVSVLSGQL-----SEKEAALTKI-----1425
Qy 234 RADGKRGSWRTDKTEARNEDEMYKIL-----LNDYEVROKQILMENAELK-----278
Db 1426 -----QTEIIEQEDLIKALHTQLEMQAKEHDERIKQLQVELCEHKKQKPEETGE 1473
Qy 279 --KVLQMKKEMISLLSPK-----KKPRERAEDGTGTV-----AISDIED--DSGSL 323
Db 1474 ESRKQIQIORKLQAALISRKALKENKSLQBELSLARGTIERLTKSLADVESQVSAONKE 1533
Qy 324 RDSVWG-----LSCOTVR-----FQL 339
Db 1534 KDTVLGRLLALQEBRDKLITEMDRSLLENQSLSSCESLKLALGLTBDKELVKEIESL 1593
Qy 340 TNS-----IRKQWRIILKSHVEKLDNOASKVHS--EGLNEE-----DVTISR 377
Db 1594 KSSKIARSTEQEKHKLQKSEYIELLSQSYENVSEARIQHVAVROEKQELCYKLKST 1653
Qy 378 QDHEQETEK----LELEIERCKEMIK-----AQQLLQOQ-----LATTC 413
Db 1654 EANKKETEKQLQEAQEMEMKEMKRFKAKSKQKQKILEEBENDRLRAEVHPAGDTAKEC 1713
Qy 414 DDDTTSLLRDYLLLEEKERLKEWTLFKQCKNFERERSFTAAIRLGLERKAFEEERA 473
Db 1714 ME--TLSSNASMKEELERVKMEYETLSKFQSLMSEKDSUSEVEVDL---KHQIEDNV 1767
Qy 474 SWVKQQLNMTNFDHNSENVK---LFSAFSGSDPDNLI VHSRPRQKKLHSLVANGVPAC 530
Db 1768 S--KQANLEATE-KHDQTNVTEEGTQSI PGTESEQDSLMSSTR-----PTC 1811
Qy 531 TSKLTKSLPASPTS-DPROTHSCVSBHSSIS-----VLNTPBESKPSSEVAR 577
Db 1812 SESVPSAKSANPAVSKDF-----SSHDEINNYLQQTIDQLKRIAGLEEBEKQKNKFSQ 1864
Qy 578 ESTQCKWSVQSRPSSREG 595
Db 1865 TLENEKNTLSQISTKDG 1882
RESULT 4
US-09-248-796A-16333
; Sequence 16333, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keitch Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208

```
; SEQ ID NO 16333
; LENGTH: 905
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-248-796A-16333

Query Match
 7.0%; Score 222.5; DB 2; Length 905;
Best Local Similarity 18.1%; Pred. No. 2e-08;
Matches 137; Conservative 140; Mismatches 264; Indels 217; Gaps 27;

QY 10 PVLCTENKNLSQYTSMTKSPSSLYSQVLCSSVPLSKNVHGVGV-FCVTGEN-----IE 63
Db 143 PIL-----LDAYKIEKKWEIIRIFPKVLQENTPVSSSEGLVLYNGLNKTCLN 196

QY 64 QTSYSLDQELTTGFPSPLYEESKEAKRELNIVAVLNCMNEILLVLR-KNLLAQ----- 117
Db 197 GEIRKLTQKNTVNFIV-----KLESNKELN-----NDVKRLDTMKEKEVNGKIDSM 247

QY 118 ---ESVETONKLGSDMDHLQCYAKLKEQLETSRREMIGLQERDRQCKNRSLSHQLL 173
Db 248 SELEIINNQDKLTDKDK-----EKIESLKLEITDRDNKIKDLESNTSSNSKEL 300

QY 174 KNEKDEVQKL-----QNIASRATQVNHVDVKERE 204
Db 301 NTELEPKEVPTKVNQDKSSSGEENTDKDSSVGDSSDSDVLTNTEISQLKSQLSKETE 360

QY 205 YNKLKERLHQLVNNKDKN-----IAMDVLNYYGRADGKRGSWRTDKTEARNEDEMY 256
Db 361 VEELTNEVRTLSQLNDKNEEIEDRLVKEIGNELVTSKDEIKSLKNSQKSTDNEDSTT 420

QY 257 K-----ILLADYEVROQIILMENAELKVLQOMKEMISLLSPOKKPRERAED----- 305
Db 421 KEDTNTQINDWELVNSQLKENKELKELDVATKE-----SKEKEKERRKSEDLQRKQIQ 475

QY 306 -----GTGTVAISDIEDDSGEL-----SRDSVWGLSCDTVRBQLTN 341
Db 476 TLKQKLESTENKESKTKDKLKNLLEKEKELEKREISLSKFKNSDSSLKLEISLSKSLTN 535

QY 342 -----SIRKQWRLKSHVEKLDNQASKVHSEGL-----NEEDVIRSQD----- 379
Db 536 KONSINELKTQDELQOQNSKNSKVEELS KSNELQSNMDFLKDKNELLTKQEVLMEN 595

QY 380 -----HEQETEKLELETERCKE-----MIKAQOQ 403
Db 596 TKSLLNQVTKLQOQEKQDVITELEKTKNKLDIVIADKSQANDMLSYKKOHEELMMKSKY 655

QY 404 LQQQLATTCDDDTTSLLRDCYLLBEKER-----LKEEWTLFKEQKNFERERRSFT 455
Db 656 SLRIE---SLEDDLTEARN---LQERTRETSNMRLLVDAEMLKQOQKDSKLE----- 704

QY 456 EAAIRLGLERKA-FEERASWVKQFLNMTNFDH-----QNSENVKLFSAFGSSD 505
Db 705 ---IARALDKAERIENNASLIRKQRELDKESVQEVKLKVEKLEN-KVMTLEKEEKS 760

QY 506 PNLIVHSRPROKHLHSVANGVACTSKLTSLPASPSTSDRQTHSCVSEHSSI----- 560
Db 761 KESLSTDNQOMSK---ELSSSTIETLALNNSTNKTIRDLNYYNNNNKLSLSEDTLRFERL 818

QY 561 ----SVLNTTPESKPEVARESTQKWSVQSRPSSRE 594
Db 819 SKNYKILNQOQYQVMKERRKGSSESTKQDKPVEATPENSD 856

RESULT 5
US-09-949-016-11433
; Sequence 11433, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11433
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-11433

Query Match
 7.0%; Score 222.5; DB 2; Length 2733;
Best Local Similarity 18.9%; Pred. No. 9.5e-08;
Matches 139; Conservative 147; Mismatches 217; Indels 233; Gaps 32;

QY 15 ENKNLSQYTSMTKMS-----PSSLYSQVLCSSVP-LSKNVHGVGVFCTGENIEQ 64
Db 699 ENENIGDQLRQLQIQVRESIDGKLPTST--DQESCSTPGLEPLFKATEQHHTQPVLES 756

QY 65 SI-----SYLDQELTTFGPPSLYE-----ESKSKEAKRELNIVAVLNCMNEILLVLRKNLLA 116
Db 757 NLCPMPWPSHSDASALQGGTSVAQIKAQLEAEKVELELVKVSSTTSELTKKSEEVFQL 816

QY 117 QBSVETQNLKLAGS--DMDHLQCYAK-LKEQLETSRREMIGLQERDRQCKNRSLSHQLL 173
Db 817 QOQINKQGLEIESLTKVTSHEAEVHAESLQKLESSQLQIAGL-EHURELOPK----- 867

QY 174 KNEKDEVQKLQNIASRATQVNHVDVKEREYNKLERLHQLVMMNKDKNIAMDVLNYYVG 233
Db 868 ---LDELQKL-----ISKKEEDVSLSGQL-----SEKEAALTKE----- 899

QY 234 RADGKRGSWRTDKTEARNEDEMYKIL-----LADYEVROQIILMENAELK----- 278
Db 900 -----QTEIIEQEDLTKALHTOLENQAQKEHDERIKQLQVELCEMKQKPEEIGE 947

QY 279 --KVLQOMKEMISLLSPQK-----KKPRERAEDGTGV-----AISDIED--DSGELS 323
Db 948 ESRAKQOQTKLQOALISRKEALKENKSLQELSLARGTIERLTKSLADVESQVSAONKE 1007

QY 324 RDSVNG-----LSCDTVVR-----EQL 339
Db 1008 KDTVLGRLLALQBERDKLITEMDRSLLLENQSLSSCSLSKALEGLTEDEKLVKEIESL 1067

QY 340 TNS-----IRKQWRLKSHVEKLDNQASKVHS---EGLNEE-----DVISR 377
Db 1068 KSSKIAETEMQEKHKEQKEYEILLQSYENVSNBAERIQHVVEAVRQEBKQELYGKLRST 1127

QY 378 QDHEQETEK---LELEIERCKEMIK-----AQQLLQOQ-----LATTC 413
Db 1128 EANKKETEKQLEAQEQEEMKEKMKRFAKSKQKILEEENDRLRAEVHAGDTAKEC 1187

QY 414 DDDTTSLLRDCYLLBEKERLKEEWTLFKEQKNFERERRSFTFAAIRGLERKAFEEPERA 473
Db 1188 ME--TLSSNASMKEELERVKMEYETLSKKFQSLMSEKDSLSEEVQDLKHQIEGNVSKQA 1245

QY 474 SW-VKQOFLNMTNFDHNSENVKLFSAFGSSDDPNLIVHSRPROKHLHSVANGVACTS 532
Db 1246 NLEATEKHDNQTNTVEGTQSIP-----GETEEQDSLMSMSTR-----PTCSE 1287

QY 533 KLTKSLPASPSTS-DRQTHSCVSEHSSIS-----VLNITPEESKPSPEVARES 579
Db 1288 SVPSAKSANPAVKDF-----SSHDINNYYLQIQDKERIAGLEBEKQKNKEFSQTL 1340

QY 580 TDQKWSVQSRPSSREG 595
Db 1341 ENENKTLTSLQISTKDG 1356

RESULT 6
```

```
US-09-949-016-6899
; Sequence 6899, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6899
; LENGTH: 1780
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6899

Query Match          7.0%; Score 220.5; DB 2; Length 1780;
Best Local Similarity 20.6%; Pred. No. 7.4e-08;
Matches 156; Conservative 117; Mismatches 235; Indels 251; Gaps 36;

QY 16 NKNLSQYTTSETKMSPPSLYS-----QQVLCSSVPLSKNVHGVFGVCTGENIEQSI 66
DB 996 SKQVKEYRIQEPNRENSPHSSIEAIWEECKEIVKAS---SKKSHQI-----EELEQOI 1045

QY 67 SYLDQELTTFFGPPSLVEESKSEAKRELNIIVAVLNCMNEILLVLRKNLLAQES---VET 122
DB 1046 EKLOAEVK--GYKDENNRLKEKEHKNQDILLK-----EKETLIQQLKEELQKNVTLDVQI 1099

QY 123 QNLKLG-----SDMDHLQSCY-AKLKEQ---LETSR-----REMIQLQERD 159
DB 1100 QHVVEGKRALSELTOGVTCYKAKIKELETTILETKQVRSKSAKLEQDILEKESI-ILKLE 1158

QY 160 RLOCKNRSLSHOLLKNEKD-----QQVLCSSVPLSKNVHGVFGVCTGENIEQSI 66
DB 1159 RNLKEFQEHLODSVKNTKDLNVKELKKEEITQLTNNLQDMKHLQLKKEEBEETNRQETE 1218

QY 182 KL-QNIIASRATQYN--HDVKKEREYNKLERLHQLVMNKKDKNIAMDVLNVYGRADGK 238
DB 1219 KLKEELSASSARTQNLKADLQKKEEDYADLKEKLT--AKKOIKQVQKEV-----1266

QY 239 RGSWRDTKTEARNEDEMYKILLNDYEQROQILMENAELKKVLOQMKKEMISLLSPQKKK 298
DB 1267 -----SVMRDEKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQ 1318

QY 299 PRERA-----EDGTGTVATSDIEDSGELSRDSVWGLSCDTVREQLTNSIRKQW 347
DB 1319 QYERACKDLNVKEKIEEDMRMTLE---EQQTQVEQDQVLEAKLEEV-ERLATELEK-W 1372

QY 348 RILKSHVEKLDNOAKSVHSEGLNEEDVISR---ODHEQETKEKLETERCKEMIKAOQQ 403
DB 1373 KEKCNLDLETKNNQRNSKEHE--NNTDVLGKLTNLQDELQESQ-KYNADR-KKWLEEKQM 1428

QY 404 LQQQLATTCCDDTTSLLRDCYLLKEERLKEEWTLFKKEQKNFRERRSF-----454
DB 1429 LITQ-----AKEAENIRN-----KEMKYAEDRRERFPQKNQMEIL 1464

QY 455 TEAATRLGLERKAFREER-----ASWVKQOFLNMTNFDHON-----SENKLFSAF 500
DB 1465 TAOQTEKOSDLQKWEERDQVAALEIQLKALISSNVQKNEIEQLKRIISSETSIETQI 1524

QY 501 S-----GSSPDNLIVHSRPRQKKLHSAVANG-----527
DB 1525 MDIKPKRISSADPKL--QTEPLSTSFISRNKIEDGSSVVLDSCEVSTENDOSTRFPKPE 1582

QY 528 -----PACTSKLTKSLPASPSSTDPRQTHSCVSEHSSI---SVLNTIP 567
```

```
US-09-949-016-7880
; Sequence 7880, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7880
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7880

Query Match          7.0%; Score 220.5; DB 2; Length 1786;
Best Local Similarity 20.6%; Pred. No. 7.5e-08;
Matches 156; Conservative 117; Mismatches 235; Indels 251; Gaps 36;

QY 16 NKNLSQYTTSETKMSPPSLYS-----QQVLCSSVPLSKNVHGVFGVCTGENIEQSI 66
DB 1002 SKQVKEYRIQEPNRENSPHSSIEAIWEECKEIVKAS---SKKSHQI-----EELEQOI 1051

QY 67 SYLDQELTTFFGPPSLVEESKSEAKRELNIIVAVLNCMNEILLVLRKNLLAQES---VET 122
DB 1052 EKLOAEVK--GYKDENNRLKEKEHKNQDILLK-----EKETLIQQLKEELQKNVTLDVQI 1105

QY 123 QNLKLG-----SDMDHLQSCY-AKLKEQ---LETSR-----REMIQLQERD 159
DB 1106 QHVVEGKRALSELTOGVTCYKAKIKELETTILETKQVRSKSAKLEQDILEKESI-ILKLE 1164

QY 160 RLOCKNRSLSHOLLKNEKD-----QQVLCSSVPLSKNVHGVFGVCTGENIEQSI 66
DB 1165 RNLKEFQEHLODSVKNTKDLNVKELKKEEITQLTNNLQDMKHLQLKKEEBEETNRQETE 1224

QY 182 KL-QNIIASRATQYN--HDVKKEREYNKLERLHQLVMNKKDKNIAMDVLNVYGRADGK 238
DB 1225 KLKEELSASSARTQNLKADLQKKEEDYADLKEKLT--AKKOIKQVQKEV-----1272

QY 239 RGSWRDTKTEARNEDEMYKILLNDYEQROQILMENAELKKVLOQMKKEMISLLSPQKKK 298
DB 1273 -----SVMRDEKLLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEEAIQ 1324

QY 299 PRERA-----EDGTGTVATSDIEDSGELSRDSVWGLSCDTVREQLTNSIRKQW 347
DB 1325 QYERACKDLNVKEKIEEDMRMTLE---EQQTQVEQDQVLEAKLEEV-ERLATELEK-W 1378

QY 348 RILKSHVEKLDNOAKSVHSEGLNEEDVISR---ODHEQETKEKLETERCKEMIKAOQQ 403
DB 1379 KEKCNLDLETKNNQRNSKEHE--NNTDVLGKLTNLQDELQESQ-KYNADR-KKWLEEKQM 1434

QY 404 LQQQLATTCCDDTTSLLRDCYLLKEERLKEEWTLFKKEQKNFRERRSF-----454
DB 1435 LITQ-----AKEAENIRN-----KEMKYAEDRRERFPQKNQMEIL 1470
```

```
QY 455 TEAAIRGLERKAFEEER-----ASWVKQOFLNMTNFDHON-----SENKLFSAF 500
Db 1471 TAQTEKSDLOKWRERDOLVALEIQLKALISSNVOKNEIEOLKRIISETSIETQI 1530
QY 501 S-----GSSDPNLIHVRPRQKHLHSHVANGV-----527
Db 1531 MDIKPKRISSADPKL--QTEPLSTSPFISRNKIEDGSSVVLDSCEVSTENDQSTRPKPE 1588
QY 528 -----PACTSKLTKSLPASPSTSDPROTHSCVSEHSSI-----SVLNTIP 567
Db 1589 LEIQFTPLQPNKMAVKGPGCTTPVTVKIPKA-----RKRKSNEMBEDLVKCNKKNATP 1642
QY 568 EESKP---SEVARESTDOKSVOSRPSREGCYSGCSSA 603
Db 1643 RTNLKPPISDRNSSVVKQKVAIRPSSKK-TYSLRSQA 1680

RESULT 8
US-09-919-172-29
; Sequence 29, Application US/09919172
; Patent No. 6673545
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 29
; LENGTH: 2125
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673545 3774181CD1
US-09-919-172-29

Query Match 6.8%; Score 215.5; DB 2; Length 2125;
Best Local Similarity 18.9%; Pred. No. 2.3e-07;
Matches 142; Conservative 136; Mismatches 260; Indels 213; Gaps 31;

QY 1 MGDWMTVDPLCTENKNLSQYSETKMGPSLSYQQVLCSSVPLSKNVHGVFGVCTGE 60
Db 452 LDDWI---QQVETTKIOENQENSKTLATOLNQOKMLVSEIEMKQSKWDE---CQKY 504
QY 61 NIEQSIYLDQELTTGFPSPLYEESKSEAKR-----ELNIVAVLNCM 103
Db 505 AEQYSATVKDYELQNTYRAMVDSQOKSPVKRRRMQSSADLIIOEFMDLRTYTLVTLM 564
QY 104 NELL-----VLQKNLLAQESVETONKLGSDMDHL 134
Db 565 TQYKFPAGSLKLEEEETKRCETSEHGAYSDDLQ-----QKATVLENSKLTGKISEL 619
QY 135 QSCYAKLKEQLETSRREMIGLOE-RDRQLQCKNRSLLHLLKNEKDEVOQLNIIA-SRAT 192
Db 620 ERWAVELKKQKSVEBELPKVREAAENELRKQQRNVEDI-----SLQIKRAESEAK 670
QY 193 QYNHD-----VKRERYNKLKERLHQLVNNKDKKNIAMD--VLNVYGRADGKRGSRWTDK 246
Db 671 QYRRELETTVREKEAAREL-ERVROLTTIEAKRAAAVEENLLNFRNQLEENTFTRRITLE 729
QY 247 TEARNEDENVKILLNDYEVROQIOLME-----NAELKKVLQOMKKEMI-----289
Db 730 DHLKRXD-----LSLNDLE--QQRNKLMEELRRKRDNEEELKLIKQEKDLAFQKQVAKQ 784
QY 290 -----SLLSPQKKKPREAED-----305
Db 785 LKEKOKIELEARRKITEIQYTCRENALPVCIPITQATSCRAVTGLQOEHDKQKAEELKQV 844
```

```
QY 306 GTGTVAISDIEDSGELSRDSVWGLSCDVTREOLTNISIRKQWRILKSHVEKLN-----359
Db 845 DELTAANRRKAEQDMRELTYE-----LNALQLEKTSSEKA-RLDKDLDETNTLRLCKL 897
QY 360 -----QASKVHSEGLNE-----EDVIS-----RODHEQETEKLELE--390
Db 898 LELERKQDAEKGYSOQLRELGRQLNQTGKAEEMQEASDLKKIKRNYQJELSLSNHEKG 957
QY 391 -----IERCKEMIKAAQQLLOQLATTCCDDDTTSLLRDCYLLBEKERLKEEWTLFKE 442
Db 958 KLQREVDRITRAHAVAENKNIQHLNSQIHSPRDEKELERLIQICQ--RKSDHLKEQFEKSHE 1015
QY 443 Q-KNPERERRSFEAAIRGLERKAFEEERASWVKQOFLNMTNFDHONSENKLFSAFS 501
Db 1016 OLLQNIKAEBEN-NDKIQRLNBELEK-SNECAEMLKQKVEELT---RONNETKLMQRIQ 1070
QY 502 GSSDPNLIHVRPRQKHLHSHV---ANGVPACTSKLTKSLPASPST-SDPROTHSCVSEH 557
Db 1071 AES--ENIVLEKQTIQORCEALKIQADGFKDQLRSTNEHLHKKOTKTEQDPQRIKCLUE- 1127
QY 558 SSISVLNITPEESKPSSEVARESTDQKWSVQS 588
Db 1128 -----DLAKSQNLVSEPKQKCDQQNIION 1152

RESULT 9
US-09-538-092-1252
; Sequence 1252, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1252
; LENGTH: 2663
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q02224
US-09-538-092-1252

Query Match 6.7%; Score 212.5; DB 2; Length 2663;
Best Local Similarity 22.4%; Pred. No. 5.4e-07;
Matches 141; Conservative 114; Mismatches 225; Indels 149; Gaps 30;

QY 17 KNLQYTSSTKSPSSLSYQQVLCSSVPLSKNVHGVFGVCTGENIEQSIYLDQELTTF 76
Db 971 ESLKQH-QETINTLASKISEEV-----SRNLHMEENTGETKDFEQKMGVGIKK---1018
QY 77 GPSPLYESKSEAKRELNIVAVLNCMNELLVLQKRNIL-LAOSVETQNLKLGSDMDHLQ 135
Db 1019 -----QDLKAKNTQTLTADVK-DNEIIBQQRKIFSLIOEKNELOQM-----LE 1060
QY 136 SCYAKLKEQLETSRREMIGLOERDQLOCKNRSLLHLLKNEKDEVOQLNIIASRATQV 195
Db 1061 SVIAB-KEQLKTDLKENI-----EMTINQOEELRL--GDELKKQOEIVAQ---EQN 1106
QY 196 HDVKKEREYNKLERLHQLVNNKDKKNIAM-----DVLNVYGRADGKRGSRWTDKTEAR 250
Db 1107 HAIK-KEGELSFTCDRLAEVBEKESQLOEKQOQLLV-----QEEEMSEHQ 1154
QY 251 ---NEDEMYKILLNDYEVROQIOLMENAEELKKVLQOMKKEMISLLSPQKKPREAEDGT 307
```



```
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
; US-09-595-684B-31

Query Match          6.7%; Score 211.5; DB 2; Length 2662;
Best Local Similarity 22.1%; Pred. No. 6.4e-07;
Matches 141; Conservative 116; Mismatches 233; Indels 147; Gaps 30;

QY 14 TENKLSQYTSKMPSSLYSQVLCS-----SVPLSKNVHGVFGVCTGENIEQSIYS 68
Db 955 TVANNIDT-QEQLRNALESKHQETINTLKSISEEVSRNLHWEENTGETKDFQKQWVG 1013
QY 69 LDELITFGPPSIYESKSEAKRELNIIVAVLNCMELLVLQRNL-LAQESVETQNLKL 127
Db 1014 IDKK-----QDLKANTQTLTADV-K-DNEIIIEQQRKIFSLIQEKNELQOM-- 1057
QY 128 GSDMDHLQCYAKLKEQLETSRREMIGLQERDQLQCKNRSLLHLLKNEKDEVQKLQNI 187
Db 1058 -----LESVIAE-KEQLKTDLKENI-----EMTIENQEEURLL---GDELKQOEIV 1100
QY 188 ASRATQYNDHVKRREYNKRLKRLHQLVMNKKDKNIAM-----DVLNYYVGRADGKRG 242
Db 1101 AQ---EKNHAIK-KEGELSRTCDRLAEVVEKLEKESQQLQEQQLNV-----1145
QY 243 RTDKTEAR---NEDEMYKILLNDYERQKQILMENAEKLVLOOMKKEMISLLSPQKKP 299
Db 1146 QEEMSEMOKKINIEINKELKNELKELTLEHMETERLELAQKLNENYEEVKSITKERK--- 1202
QY 300 RERAEDGTGVAISDIEDSGELSRDSVNGLSCTVREQLTNSIRKQWRI-----LKSH 353
Db 1203 -----VLKELQ-KSFETERDHLRGY-----IREIATGLQTKKEELKIAHHLKHEH 1246
QY 354 VEKLDNQASKVHSEGLNEEDVSRQDHEQTEKLELEI-----ERCKEMIKAQOQ 403
Db 1247 QETIDELRSVSEK---TAQINTQDLEKSHYKLOEEIPVLHHEEQELLNVKKVSTQET 1303
QY 404 LIQOQLAT--TDDDTTSLRLDCYLLEEKRLKEWTLFKQKKNFERRRSFTAAIRL 461
Db 1304 MNELELLTQSTTKDSTTLAR---IEMERLRLNEKFOESQOEIKSLTKERDNL--KTIKE 1358
QY 462 GLE-----RKAFESERASWKO-QFLANNTNFDHONSENKVLFSAP-----500
Db 1359 ALEVHDQLKEHIRETLAKIQESQSQOSQSLNNKEDNTTKIVSMEQFKPKDSALLRI 1418
QY 501 -----SGSSDPDNLIVHSRPRQKKLHVSANGVPACTSKLTKLPAS--PSTSD 547
Db 1419 EIEMLGLSKRLQESHDEMSVAKEKDDLQRLQEVLSQESDQLKENIKEIVAKHLETEBEL 1478
QY 548 ROTHSCVSE-HSISVNLNTPESKPSVARESTDQK 583
Db 1479 KVAHCLKEQEETINELRVNLSE-KETEI---STIQK 1511
```

RESULT 12

```
US-08-685-871-2
; Sequence 2, Application US/08685871
; Patent No. 6013499
; GENERAL INFORMATION:
; APPLICANT: NARUMIYA, Shuh
; APPLICANT: INAMATSU, Akihito
; TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,871
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184102
; FILING DATE: 25-JUN-1996
; APPLICATION NUMBER: JP 7-262553
; FILING DATE: 14-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/845
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-685-871-2

Query Match          6.7%; Score 210.5; DB 2; Length 1354;
Best Local Similarity 21.7%; Pred. No. 3e-07;
Matches 134; Conservative 103; Mismatches 219; Indels 161; Gaps 26;

QY 83 EESKSEAKRELNIIVAVLNCMELLVLQRNLLAQESVETQNLKSGSDMDHLQSCYAKLK 142
Db 570 ENSKS-----TSQLESINR--ELQERNRI-----LENSKQSTDKD-----YYQLQ 607
QY 143 EOLETSSRR-----EMIGLOERDQLQCKNRSLLHLLKNEKDEVQKLQNIASRATQY 196
Db 608 AILEAERRDRGRHSEMIG-----DLQARITSLQEVVHKLHNEKVEGERKEAQDMLNH 661
QY 197 DVYRKER-----EYNKLERLHQLVMNKK-----DKNIAMDVLNVVG--RADGKRG 240
Db 662 SEKKNKLEIDLNYKLKSLQORLEQVNEHKVTKARLTDKHQSIEEAKSVAMCEMKCKL 721
QY 241 SWRTDKTEARN-----EDMYKILLNDYEQY 267
Db 722 EEREAREAKENRVVQIEKQCSMLDVLQKSOQKLEHLTGKMERMEDEVKNLTILQEQESN 781
QY 268 KOILMENA-----ELKKVLOQMCKEMISLLSPQKKPRERAEDGTGTVAISDIED 318
Db 782 KRLLLQNELKTOAFEADNLKGLKQMKQKEINTLLEAKRLLEFELAQ-----LTKQYRG 835
QY 319 SGELSRDSVNGLSCTVREQLTNSIRK-OWRILKSHVEKLDNQASKVHSEGLNEEDVISR 377
Db 836 EQQMBE-----LQDQLEAEQVFTLYKTQVKELKEEIEKNRENKLTQELQONKETLAT 890
QY 378 Q-----DHEQETEKL-----ELETERCKEMIKAQOQLLQOQLATTCDDDTTSL--R 422
Db 891 OLDLAEAKESQALARGLEBQYFELTQESKKAASRNQEIITDKOHTVSRLEANSMLTK 950
QY 423 DCYLL-----EKEKRLKEWTLFKEQ-----KKNFERRRSFTAAIRIG--L 463
Db 951 DIEILRRENEELTEKMKKAEYKLEKEEISNLKAAFEKINTERTLTQAVNKLAEIM 1010
QY 464 ERKAFEEERASWVKQFLNMTNFDHONSENKVLFSAFSGSSDPDNLIVHSRPRQKKLHVS 523
Db 1011 NRKDFKIDRKK-----ANTQDLRKKCKENRKLQELNQEOKFKNQWVK--HQELNDM 1062
QY 524 -ANGVPACTSKLTKSLPASPTSDPROTHSCVSEHSSI SVLNTITPEESKPSVARESTDQ 582
Db 1063 QALVEECAHRELQMLQASKESDIEQLRA-----KLDDLSDSTSVASFPSEADTD- 1113
```

Qy 583 KWSVQSPRSSR-EGCYS 598
Db 1114 ----GNLPESRIEGWLS 1126

RESULT 13
US-08-685-576-4
; Sequence 4, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kozo
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-685-576-4

Query Match 6.6%; Score 208.5; DB 1; Length 1388;
Best Local Similarity 20.1%; Pred. No. 4.4e-07;
Matches 109; Conservative 109; Mismatches 186; Indels 139; Gaps 21;

Qy 67 SYLDELTTFGPSPSYESKSKAKRELINIVAVLNCMNEILLVLRKLLAQAESVETONLK 126
Db 402 AFVGNQLPFIQF-TYREN-----LLSDSPSCRENDSIQSRKN---BESQIQK-K 448

Qy 127 LGSMDHLQSCYAKLKEQLTSRRREM-IGLQERDRQLQ---CKNRSLLHQLLNKNEDEVOK 182
Db 449 LYTLEEHL-SNEMOAKEEQLCKSVNTRLEKTAKLEBEEITLRSVESALRQLERKAL 507

Qy 183 LQNTIASRATQYNHDKVRK---EREYNKLERLHOLVNNKKDKNTAMDVLNVVGR-ADGK 238
Db 449 LYTLEEHL-SNEMOAKEEQLCKSVNTRLEKTAKLEBEEITLRSVESALRQLERKAL 507

Qy 183 LQNTIASRATQYNHDKVRK---EREYNKLERLHOLVNNKKDKNTAMDVLNVVGR-ADGK 238
Db 508 LQHKNAEYQKADHEADKRNLENDVNSLKQLEDLKKRNQNSQISTEKNVQLQRLDET 567

Qy 239 RGSWRD-----KTEARNEDEMYKILLNDYEVROKQIILMENAEKLVLOQMKKEMIS 290
Db 568 NALLRTESDTAARLRKTOAESSKQIQLESNNRDLQDNCLLETAKLK-----LEKEFIN 622

Qy 239 RGSWRD-----KTEARNEDEMYKILLNDYEVROKQIILMENAEKLVLOQMKKEMIS 290
Db 568 NALLRTESDTAARLRKTOAESSKQIQLESNNRDLQDNCLLETAKLK-----LEKEFIN 622

Qy 291 LLSPOKKKPRERA-----EDGTGTVA-----ISDIEDD 318
Db 623 LQSALESERRDRTHGSEIINDLQGRICGLEEDLKNGKILLAKVELEKRLQLOERFTDLEKE 682

Qy 319 SGELSRDSVWGLSC-----DTVREQLTNSIRK-----Q 346
Db 683 KSNMEIDMTYQLKVIQOSLEQEAHAKTKARLADKNKIYESIEEAKSEMKEKKLLE 742

Qy 347 WRILKSHVEKLDNOASKVHS-----EGLNEEDVISRQDH-----BQETEK 386
Db 743 ERTLKQKVENLLLEAEKRCSLDDCDLKOSQKINELLKQKDVNLNEDVRNLTUKIQQETQK 802

Qy 387 -----LELEIERCKEMIKAAQQLLQ-----QLATTCDDDTTSLLR-----DCYLLEE 429
Db 803 RCLTQNDLKMQTQVNTLKMSEKQLKQENNHLMEMKMNLEKQNAELRKERQDADQMKEL 862

Qy 430 KERLKEW---TLFEKQCKNFERERRSFTEAAIRLGLERKAFEEERASAMVQOFLNMTNF 486
Db 863 QDQLEAEQYFTLYKTQVRELKECECEKTLKGLKQKQQLQDQERDLSLAAQLTITLTKA 922

Qy 487 DHQ 489
Db 923 DSE 925

RESULT 14
US-09-976-594-296
; Sequence 296, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 296
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 064987CD1
US-09-976-594-296

Query Match 6.6%; Score 208.5; DB 2; Length 1388;
Best Local Similarity 20.1%; Pred. No. 4.4e-07;
Matches 109; Conservative 109; Mismatches 186; Indels 139; Gaps 21;

Qy 67 SYLDELTTFGPSPSYESKSKAKRELINIVAVLNCMNEILLVLRKLLAQAESVETONLK 126
Db 402 AFVGNQLPFIQF-TYREN-----LLSDSPSCRENDSIQSRKN---BESQIQK-K 448

Qy 127 LGSMDHLQSCYAKLKEQLTSRRREM-IGLQERDRQLQ---CKNRSLLHQLLNKNEDEVOK 182
Db 449 LYTLEEHL-SNEMOAKEEQLCKSVNTRLEKTAKLEBEEITLRSVESALRQLERKAL 507

Qy 183 LQNTIASRATQYNHDKVRK---EREYNKLERLHOLVNNKKDKNTAMDVLNVVGR-ADGK 238
Db 508 LQHKNAEYQKADHEADKRNLENDVNSLKQLEDLKKRNQNSQISTEKNVQLQRLDET 567

Qy 239 RGSWRD-----KTEARNEDEMYKILLNDYEVROKQIILMENAEKLVLOQMKKEMIS 290
Db 568 NALLRTESDTAARLRKTOAESSKQIQLESNNRDLQDNCLLETAKLK-----LEKEFIN 622

```
QY 291 LLSPOKKKPERA-----EDGTGTV-----ISDIEDD 318
Db 623 LQALSERDRTHGSEIINDLQRIICGLBEEDLKGKILLAKVELEKRLQERFTDLEKE 682
QY 319 SGELSRDSVWGLSC-----DTVREQLTNSIRK-----Q 346
Db 683 KSNMEIDMTYQKVIQOSLEQEEAEHAKTKARLADKKNKIYESIEEAKSEAMKEMEKLLJE 742
QY 347 WRILKSHVEKLDNQASKVHS-----EGLNEEDVISRODH-----BOETEK 386
Db 743 ERTLKQKVENLLLEAEKRCSLDCDLKQSOOKINELLKQKDLVNEVRNLTLKIEQETOK 802
QY 387 -----LELEIERCKEMIKAOQILQO-----QLATTCDDDTTSLR-----DCVLLLE 429
Db 803 CCUTQNDLKNQTOQVNTLKNSEKQKLOENHLMEMKQNLKQNAELRKERQDADGQMKEL 862
QY 430 KERLKEW---TLFKEQKKNFERERSFTEAAIRLGLERKAFEEERASVWVKQOFLNMTNF 486
Db 863 QDOLEAEQVFTLYKTQVRELKECEKTYLKGELQOQKQELQDERDLSLAAQLEITLTKA 922
QY 487 DHQ 489
Db 923 DSE 925
```

RESULT 15

```
US-09-538-092-1260
; Sequence 1260, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1260
; LENGTH: 2704
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q03001
US-09-538-092-1260
```

```
Query Match 6.6%; Score 208; DB 2; Length 2704;
Best Local Similarity 18.5%; Pred. No. 1.2e-06;
Matches 142; Conservative 136; Mismatches 260; Indels 228; Gaps 31;
```

```
QY 1 MGDWMTVPDLCTENKNLSQYTSFKMSPSLSYQQVLCSSVPLSKNVHGVGVFCTGE 60
Db 1016 LDDWI---QVETQTKIQENQSPENSKTLATQNLQKMLVSEIMKQSKWDE---CQKY 1068
QY 61 NIEQISYLDQLTTFGFPLSYEKSKEAKR-----ELNIVAVLNCM 103
Db 1069 AEQYSATVKDYELQMTYTRAMVDSQKSPVKRRMQSSADLIQBFMDLTRYTALVTIM 1128
QY 104 NELL-----VLRKNLLAQES 119
Db 1129 TOYIKFAGDSLRLEEEVINSHPVNIKEIKRKCTSEHGAYSDLLQR-----OKA 1183
QY 120 VETONLKLGSMDHLQSCYAKLKEQLTSRRMIGLQE--RDRQLQCKNRSRLHQLLNKXD 178
Db 1184 TVLENSKLTGKISELRMVAELKQKSRVEELPKVREAENELRKQQRNVEDI----- 1237
```

```
QY 179 EVQKLQNIITA-SRATOYNHD-----VVKEREYVKNLKERLHQLVMNKKDKNIAMD--VLNY 231
Db 1238 ---SLQKIRAESEAKQYRELETIIVREKEAAEREL-ERVRLQLTIEAAKRAAAVEENLLNF 1293
QY 232 VGRADGKRGSRWTDKTEARNEDEMYKILLNDVEYRQKQILME-----NAECLKVLOQ 283
Db 1294 RNQLENTTFRITLEDHLKRD-----LSLNDLE-QQKNKLMBELRRKRNEEELLKLIQ 1348
QY 284 MKKEMI-----SLLSPQ 295
Db 1349 MEKDLAFQQAQVAEKQKQKIELEARRKI TEIQVTCRENALPVCPIQATSCRAVTGLQ 1408
QY 296 KKKPRERAED-----GTGTVAISDIEDDSGELSRDSVWGLSCDVTREQLTNSIRKQWRIL 350
Db 1409 QSHDKQKAEELKQOVDELTAANRKAQDMRELTYE-----LNALQLEKTSSEKA-RLL 1461
QY 351 KSHVEKLDN-----QASKVHSEGLNE-----EDVIS-----R 377
Db 1462 KDKLDETNTLRLCLLELERKQDAEKGYQQQRELGRQLNQTTGKAEAMQASDLKKIK 1521
QY 378 QDHEQTEKLELE-----IERCKEMIKAOQQLQOQLATTCCDDTTSLLRDCYLL 427
Db 1522 RNYQLELESANHEKGLQREVDRIITRAHAVAEEKNIQHLNSQIHSFDEKELERLQICQ-- 1579
QY 428 EKERLKEEWTLFKEQ-KKNFERRRSFTEAAIRLGLERKAFEEERASVWVKQOFLNMTNF 486
Db 1580 RKSDHLKEQFEKSHEQLLQNIKAEN-NDKIQRLNEELEK-SNECAEMLKQKVEBLT-- 1635
QY 487 DHONSENVKLFSAFSGSSDDPNLIVHSRPROKKLHSV---ANGVPACTSKLTKSLPASPS 543
Db 1636 -RQNETKLMQRIQES--ENIVLEKQTIQRCALKIQADGFKDQLRSTNEHLHKQTK 1692
QY 544 T-SDFRQTHSCVSEHSSISVLNITPEESKPSVARESTDQKWSVQS 588
Db 1693 TQDQFORKIKLEE-----DLAKSQNLVSEFKQKCDQDNIIQN 1731
```

Search completed: June 12, 2006, 19:16:08

Job time : 56 secs

THIS PAGE LEFT BLANK

Db 224 ATGGGAGATTGGATGACTGTTACAGATCCA-----GAAAGCAAAACTATCTCT 271
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db 272 CAATATACCTCGAAGCAAAAGATGCTCCATCAAGTTTATATCTCACAGCAAGTGCTATGT 331
Qy 41 SerSerValProLeuSerLysAenValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 332 TCTTCAATACCTTTATCGAAAGATGTCAGAGTTTTCAGTCCCTCTCGACAGAGAT 391
Qy 61 AsnIleGlnSerIleSerTyrLeuAspGlnGlnLeuThrThrPheGlyPheProSer 80
Db 392 AATATTGAACACAGATATCTCATATCTTGATCAGGAATGACTACTTTTCGTTTCCTTCA 451
Qy 81 LeuTyrGluGlnSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db 452 TTATATGAAGAATCCAAAGGTAAAGAGACAAGAGAGAGTTAAATATATAGTGTACTPA 511
Qy 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuAlaGlnGluSerVal 120
Db 512 AATTGTATGAATGAGCTGCTTGCTTCAGCGGAAGAACCTTCTAGCTCAGGAAATGTG 571
Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisGlnGlnSerCysTyrAlaLys 140
Db 572 GAGACACAGAAATTTGAAGCGCGGAAGTGATATGGACCATCTACAGAGCTGTACTCAAAA 631
Qy 141 LeuLysGlnGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgParg 160
Db 632 CTTAAGGAACAACCTGGAACCTCCAGAGGGAATGATTGGGCTTCAGGAAGAGACAGA 691
Qy 161 GlnLeuGlnCysLysAenArgSerLeuHisGlnLeuLysAsnGlnLysAspGluVal 180
Db 692 CAGTTACAAATGTAAGAACAGGAATTTGCATCAGCTACTAAAGAAATGAGAAAGATGAGGTG 751
Qy 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Db 752 CAAAAATTACAAATATCATTTGCAAGTCGAGCTACTCAGTAAATCATCATATGAAGAGA 811
Qy 201 LysGluArgGluTyrAsnLysLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db 812 AAAGACGGTGAATATTAATAACTGAAGGAACGTCTACATCAACTGTTTATGAACAGAAA 871
Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db 872 GATAAGAAATAGCTATGACATTTTGAATTAATGTCGGCAGAGCTGATGNAAGAGGC 931
Qy 241 SerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
Db 932 TCCTGGAGGACTGGTAAAACTGAAGCCAGGAATGAAGATGAATGTATAAAATTCCTTTG 991
Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysVal 280
Db 992 AATGATTATGAATATCGTCAGAAACAAATTCCTAATGGAATAATGCAAGACTTAAGAAGGTT 1051
Qy 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
Db 1052 CTTCAACAATGAAAGGAATGATTTCTCTTCTTCTCCCAAAAGAAAGAACTAGA 1111
Qy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320
Db 1112 GAAAGAGTAGTAGTAGTACAGGAAGCTGTT---ATTTCGATGTTGAAGAGATCCGGG 1168
Qy 321 GluLeuSerArgAspSerValTyrGlyLeuSerCysAspThrValArgGlnGlnLeuThr 340
Db 1169 GAACTNAGCAGAGAGAGTAGTGTGGACCTTTCTGTGAACCTGTGAGAGAGCGCTTACA 1228
Qy 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValIcLulysLeuAspAsnGln 360
Db 1229 AACAGCATCAGAAAAACAGTGGAGAAATTTTGAAGAGTCATGTAGAAAAGCTTATAACCAA 1288
Qy 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
Db 1289 GTTTCAAAGGTACACCTGGAGGTTTAAATGATGAAGATGTAAATCTCTACGACAAGACCAT 1348

Qy 381 GluGlnGluThrGluLysLeuGluLeuGluArgCysValGluMetIleLysAla 400
Db 1349 GAACAGAAACTGAAAACTCGAGTTAGAAATTCAGCAATTCAGCAAAATGATTAATACT 1408
Qy 401 GlnGlnGlnLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
Db 1409 CAGCAACAGCTTTTACACGACGAGCTCGTACTGCTATGATGATGATACCACTTCACTA 1468
Qy 421 LeuArgAspCysTyrLeuLeuGluLysGluArgLysGluGluThrThrLeuPhe 440
Db 1469 TTACGAGACTGTTATTTGTTGGAAGAAAGCAAGCTCTCAAAGAAAGAAATGTCCTCTTTT 1528
Qy 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
Db 1529 AAAGAGCAGAAAGAAAGAAATTTTGAAGGAGAGACGAGCTTTACAGAGCGCTATTGCG 1588
Qy 461 LeuGlyLeuGlu----- 464
Db 1589 CTGGGATTGGAGATTGGCTTTCTCTGCCAAGTGTTTTACACAGAAATCTCTTTTGGGT 1648
Qy 464 ----- 464
Db 1649 ATCCATCATTCATCCACGGTCAGTAGTGTGCTGTCAGAAAGACAGTGCACCAAGACTAT 1708
Qy 465 -----ArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnGln 479
Db 1709 ATTTGCTGTTTAAAAAGAAAGGCATTTTGAAGAGAAAGAGCCAGTTCGTTAAAGCAGCA 1768
Qy 479 nPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAl 499
Db 1769 GTTTCTAAATATGACTACTCTTTGACCACCAAGAACTCAGAAAAATGTGAAACTTTTTCAGTGC 1828
Qy 499 aPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLys 519
Db 1829 CTTCTCAGAGAGTTCTGATTGGGACAACTCTATATAGTGCACCTCGAGCGCAGCCGCAAGAA 1888
Qy 519 sLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuPr 539
Db 1889 GCCTCACAGTGTGCTAATGGGTCTCCAGTTTGCATGTCTAACTTACTTAATCTCTTCC 1948
Qy 539 oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe 559
Db 1949 TGCTTCACTTCCACTTCAGACTTTTGCACAGACAGCTTCTGTCATATCTGAACATAGTTC 2008
Qy 559 rLisSerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSe 579
Db 2009 AATCAATGTACTGAATATACTGCTGAAGAAATTAACCAAAATCAGGTTGAGGAGAACG 2068
Qy 579 rThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGln 599
Db 2069 TACAAATCAAAATGGAGTGTGGCGTCAAGACCTGGATCAGGGAAGGTTGCTATAGTGG 2128
Qy 599 yCysSerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
Db 2129 ATGCTCCTTGAGCTACACAAATTTCTCATGTAGAAAAAGATGACTTACCT 2177

RESULT 2

US-09-513-999C-3004
; Sequence 3004, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59 US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681

1 SEQ ID NO 5562
2 LENGTH: 8590
3 TYPE: DNA
4 ORGANISM: Human
5 US-09-949-016-5562

Alignment Scores:

Prod. No.: 1-77e-12 Length: 8590
Score: 222.50 Matches: 139
Percent Similarity: 38.9% Conservatives: 147
Best Local Similarity: 18.9% Mismatches: 217
Query Match: 7.0% Indels: 233
DB: 3 Gaps: 32

US-10-644-084-2 (1-615) x US-09-949-016-5562 (1-8590)

QY 15 GluAsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSer----- 29
DB 2097 GAAATGAGATATTGGAGACAGCTAAGGCAACTCCAGATTCAGTAAGGGAATCCATA 2156
QY 30 -----ProSerLeuTyrSerGlnGlnValLeuCysSerValPro--- 44
DB 2157 GACGGAAAACTCCCAAGCACA-----GACCAGCAGGAATCGTGTCTCCACTCCAGGT 2210
QY 45 LeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGluAsnIleGluGln 64
DB 2211 TTAGAAGAACCTTTATTCAAGCCACAGACAGCATCACACTCAACCTGTTTAGAGTCC 2270
QY 65 SerIle-----SerTyrLeuAspGlnGluLeuThrThrPheGlyPhePro 79
DB 2271 AACTTGTGCCAGACTGGCGCTTCTCATCTGAAGATGCGAGTCTCTGCGAGGCGGAAC 2330
QY 80 SerLeuTyrGlu-----GluSerLysSerLysGluAlaIysArgGluLeuAsnIle 96
DB 2331 TCTGTGCCAGATTAAAGCCAGCTGAAGGAAATAGAGGCTGAGAAAGATAGAGTTAGAA 2390
QY 97 ValAlaValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuAla 116
DB 2391 TTGAAAGTTAGTTCTACACAACTGAGCTTACTAAAAATCAGAGAGTATTTCAGTTA 2450
QY 117 GlnGluSerValGluThrGlnAsnLeuLysLeuGlySer-----AspMetAspHisLeu 134
DB 2451 CAAGAGCAGATAATAAAGAGGGTTTGAATCGAGAGTCTAAAGACAGATATCCCATGAA 2510
QY 135 GlnSerCysTyrAlaLys---LeuLysGluGlnLeuGluThrSerArgArgGluMetIle 153
DB 2511 GCTGAAGTCCATGCCCAAGCTGCGACAGAAATGGAAGAGCAGCCCACTACAAATGCT 2570
QY 154 GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu 173
DB 2571 GGCCTA--GAACATCTAAGAGATTGCAACCTAAA----- 2603
QY 174 LysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArgAlaThrGln 193
DB 2604 -----CTGGATCAACTGCAAAAACCTC----- 2624
QY 194 TyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHis 213
DB 2625 -----ATAAGCAAAAGGAAGAGAGCTTTAGTACTCTTCTCGGACAACT--- 2669
QY 214 GlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrValGly 233
DB 2670 -----AGTGAGAAAGAGCAGCTCTCACTAAAAATA----- 2699
QY 234 ArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAsp 253
DB 2700 -----CAGACAGAGATAATAGAACCAAGAA 2723
QY 254 GluMetTyrLysIleLeu-----LeuAsnAspTyrGluTyrArg 266
DB 2724 GATTTAATTAAGCTCTGCATACACAGCTAGAAATGCAAGCCCAAGAGCATGATGAGAG 2783
QY 267 GlnLysGlnIleLeuMetGluAsnAlaGluLeuLys----- 278

DB 2784 ATAAAGCAGCTACAGGTGTAACCTTTGTCAAATGAAGCAAAAACCAAGAGAGATTGGAA 2843
QY 279 -----LysValLeuGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLys 296
DB 2844 GAAAGTAGACCAAGCAACAATACAAAGGAAACTGCAAGCTGCCCTTTATTTCCCGAAA 2903
QY 297 -----LysLysProArgGluArgAlaGluAspGlyThrGlyThrVal 310
DB 2904 GAAGCACTAAAGAAAACAAAAGTCTCCAAGGGAATTTGTCTTTGGCCAGAGTACCATT 2963
QY 311 -----AlaIleSerAspIleGluAsp-----AspSerGlyGluLeuSer 323
DB 2964 GAACGTCTCACCAAGTCTCTGCGACATGTGGAAGCCAAAGTTTCTGCTCAAAATAAGAA 3023
QY 324 ArgAspSerValTrpGly----- 329
DB 3024 AAAGATACGGTCTTAGGAAGGTTAGTCTCTTCAAGAAGAAAGAGACAAACTCATTA 3083
QY 330 -----LeuSerCysAspThrValArg 336
DB 3084 GAAATGGACAGGTCTTTATTGGAATAATCAGAGTCTCAGCAGCTCTGTGAAAGTCTAAA 3143
QY 337 -----GluGlnLeu 339
DB 3144 CTAGCTCTAGAGGCTTTACTGAGACAAGGAAAGTTAGTGAAGGAAATTCGATCTTTG 3203
QY 340 ThrAsnSer-----IleArgLys 345
DB 3204 AAATCTTTAAGATTGCAAGAAAGTACTAGTGGCAAGAGAAACAACAAGAGAGCTACAAA 3263
QY 346 GlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHis 365
DB 3264 CAGTATGAATCTCTCGCAGTCTATCAGAAATGTTAGTAATGAAGCAGAAAGGATTCAG 3323
QY 366 Ser-----GluGlyLeuAsnGluGlu-----AspValIleSerArg 377
DB 3324 CATGTGTGGAAGCTGTGAGGCAAGAGAAACAAGAACTGTATGGCAAGTTAAGAGACACA 3383
QY 378 GlnAspHisGluGlnGluThrGluLys-----LeuGluLeuGluLysGluArg 393
DB 3384 GAGGCAAAACAAGAGGAGACAGAAAGAGCAGTTGCAAGAGCTGAGCAAGAAATGGAGGAA 3443
QY 394 CysLysGluMetIleLys-----AlaGlnGlnGlnLeuGlnGln 408
DB 3444 ATGAAAGAAAGATGAGAAAGTTTGCTAAATCTAAACAGCAGAAATCTCTAGAGCTGAA 3503
QY 409 -----LeuAlaThrThrCys 413
DB 3504 GAAGAGATGACCGCTTTAGGCGAGAGGTGCACCTCGCAGGAGATACAGCTAAAGAGTGT 3563
QY 414 AspAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluLysGluArgLeu 433
DB 3564 ATGGAA-----ACACTTCTTTCTCCCAATGCCACATGAAGGAAGAACTTGAAGGGTCT 3617
QY 434 LysGluGluTrpThrLeuPheLysGluGlnLysLysAsnPheGluArgGluArgSer 453
DB 3618 AAAATGGAGTATGAACCCCTTTCTAAGAAGTTTTCAGTCTTTTAATGCTGTGAGAAGACTCT 3677
QY 454 PheThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArgAla 473
DB 3678 CTAAGTGAAGAGGTTCAAGATTTTAAAGCATCAGATAGAGGTTAATGTTATCTTAAACAAGCT 3737
QY 474 SerTrp---ValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGlu 492
DB 3738 AACCTAGAGGCCACGAGAAACATGATAACCAACCAATGTCTCCTGAAGAGGAGACACAG 3797
QY 493 AsnValLysLeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHis 512
DB 3798 TCTATACCA-----GGTCAGACTGAAGAGCAAGACTCTCTGATGATGAGC 3842
QY 513 SerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCysThrSer 532

Db 3843 ACAAGA-----CCTACATGTTTCAGAA 3863
QY 533 LysLeuThrLysSerLeuProAlaSerProSerThrSer---AspPheArgGlnThrHis 551
Db 3864 TCGGTTCCATCAGCGAAGAGTGCCAAACCCCTGCTGTAAGTAAGGATTTTC-----3911
QY 552 SerCysValSerGluHisSerSerIleSer-----561
Db 3912 -----AGTCACATGATGAATTAATACTACTACAGCAGATGATGATCAGCTCAAA 3962
QY 562 -----ValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSer 579
Db 3963 GAAAGAATTCGTGATTAGAGGAGGAGACGACGAGAAACAAAGCAATTTAGCCAGACTTTA 4022
QY 580 ThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGly 595
Db 4023 GAAATGAGAGAAATACCTTACTGAGTCAGATATCAACAAAGGATGGT 4070
RESULT 6
US-09-949-016-17304
; Sequence 17304, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17304
; LENGTH: 38575
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(38575)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17304
Alignment Scores:
Pred. No.: 2,25e-11 Length: 38575
Score: 222.50 Matches: 139
Percent Similarity: 38.9% Conservative: 147
Best Local Similarity: 18.9% Mismatches: 217
Query Match: 7.0% Indels: 233
DB: 3 Gaps: 32
US-10-644-084-2 (1-615) x US-09-949-016-17304 (1-38575)
QY 15 GluAsnLysAsnLeuSerGlnTyThrSerGluThrLysMetSer-----29
Db 4097 GAAATGGAATATTGGACACAGCTAAGGCAACTCCAGATTCAGTAAGGGAATCCATA 4156
QY 30 -----ProSerSerLeuTyThrSerGlnGlnValLeuLysSerSerValPro---44
Db 4157 GACGGAAAACTCCCAAGCACA-----GACCAGCAGGAATCGTGTCTTCCACTCCAGGT 4210
QY 45 LeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGluAsnIleGluGln 64
Db 4211 TTAGAAGACCTTTATTCAAGCCACAGACAGATCAGCTCAACCTGTTTATAGAGTCC 4270
QY 65 SerIle-----SerTyLeuAspGlnGluLeuThrThrPheGlyPhePro 79
Db 4271 AACTTGTGCCAGACGTGGCCCTTCTCATTTCTGAAGATGCGAGTGCTCTGCGAGGCGGAACT 4330

QY 80 SerLeuTyThrGlu-----GluSerLysSerLysGluAlaLysArgGluLeuAsnIle 96
Db 4331 TCTGTTGCCAGCATTAAGGCCCGAGCTGAAGAAATAAGAGCTGAGAAAGTAGAGTTAGAA 4390
QY 97 ValAlaValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuAla 116
Db 4391 TTGAAGTTAGTTCTACACAACTGAGCTTACTAAAAAATCAGAACGAGGTATTTTCAGTTA 4450
QY 117 GlnGluSerValGluThrGlnAsnLeuLysLeuGlySer-----AspMetAspHisLeu 134
Db 4451 CAAGACGACATAAATAACAGAGGTTTGAATAATCGAGAGTCTAAAGACAGATATCCCATGAA 4510
QY 135 GlnSerCysTyThrAlaLys---LeuLysGluGlnLeuGluThrSerArgArgGluMetIle 153
Db 4511 GCTGAAGTCCATGCCGGAAGCCTGCAGCAGAAATTCGNAAGCAGCAACATACAAATTGCT 4570
QY 154 GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu 173
Db 4571 GGCCTA--GAACATCTAAGAGAAATTGCAACCTTAA-----4603
QY 174 LysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArgAlaThrGln 193
Db 4604 -----CTGGATGAACTGCAAAACTC-----4624
QY 194 TyrAsnHisAspValLysArgLysGluArgGluTyThrAsnLysLeuLysGluArgLeuHis 213
Db 4625 -----ATAAGCAAAAGGAAGACGTTAGTACTCTTTCTCGACAACATTT---4669
QY 214 GlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyThrValGly 233
Db 4670 -----AGTGAGAAAGAGCAGCTCTCCTCAATAATA-----4699
QY 234 ArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAsp 253
Db 4700 -----CAGACAGAGATAATAGAACAGAA 4723
QY 254 GluMetTyThrLysIleLeu-----LeuAsnAspTyThrGluTyArg 266
Db 4724 GATTTAATTAAGGCTCTGCATACAGCTAGAAAATGCAAGCCAAAGAGCATGATGAGAGG 4783
QY 267 GlnLysGlnIleLeuMetGluAsnAlaGluLeuLys-----278
Db 4784 ATAAAGCAGCTACAGGTGGAACTTTGTGAATATGAAGCAAAAACAGACAGAGATTGAGAA 4843
QY 279 -----LysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLys 296
Db 4844 GAAAGTAGACAAAGCAACAAATACAAAGGAAACTGCAAGCTGCCCTTTATTTCCCGAAAA 4903
QY 297 -----LysLysProArgGluArgAlaGluAspGlyThrGlyThrVal 310
Db 4904 GAAGCACTAAAGAAACAAAAGTCTCCAAGAGGAATTTGTCTTTGGCCAGAGGTACCATTT 4963
QY 311 -----AlaIleSerAspIleGluAsp-----AspSerGlyGluLeuSer 323
Db 4964 GAACGTCTCACCAAGTCTCTGCAGATGTGGAAAGCAAGTTTCTGCTCAAAATAAGAA 5023
QY 324 ArgAspSerValTrpGly-----329
Db 5024 AAAGATACGGTCTTAGGAAGGTAGTCTCTTCAAGAGAAAGAGACAACTCATTACA 5083
QY 330 -----LeuSerCysAspThrValArg 336
Db 5084 GAAATGGACAGGTCTTTATTGGAAAAATCAGAGTCTCAGCAGCTCCTGTGAAAAATGTAAAA 5143
QY 337 -----GluGlnLeu 339
Db 5144 CTAGCTCTAGAGGCTTTACTGAACAGCAAGGAAAGTTAGTGAAGAAATTCGAATCTTTG 5203
QY 340 ThrAsnSer-----IleArgLys 345
Db 5204 AAATCTTCTAAGATTGCAAGAAAGTACTGAGTGGCAAGAGAAACACAGGAGCTACAAAA 5263
QY 346 GlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHis 365


```
QY 234 ArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAsp 253
Db 37278 -----CAGACAGAGATAATAGAACCAAGAA 37301
QY 254 GluMetTyrLysIleLeu-----LeuAsnAspTyrGluTyrArg 266
Db 37302 GATTTAATTAAGGCTCTGCATACACAGCTAGAAATGCAGCCAAAGAGCATGATGAGAGG 37361
QY 267 GlnLysGlnIleLeuMetGluAsnAlaGluLeuLys----- 278
Db 37362 ATAAAGCAGCTCAGGTGGAACCTTTGTGAATGAAGCAAAAAACCAAGAGAGATTGGAGAA 37421
QY 279 -----LysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLys 296
Db 37422 GAAAGTAGAACCAAGCAACAATACAAAGAACTGCAAGCTGCCCTTATTTCCCGAAAA 37481
QY 297 -----LysLysProArgGluArgAlaGluAspGlyThrGlyThrVal 310
Db 37482 GAAGCACTAAAGAAACAAAGTCTCCAAGAGGAATTGCTTTGGCCAGAGTACCATT 37541
QY 311 -----AlaIleSerAspIleGluAsp-----AspSerGlyGluLeuSer 323
Db 37542 GAACGTCTCACCAAGTCTCTGGCAGATGTGGAAGCAAGTTTCTGCTCAAAATAAGAA 37601
QY 324 ArgAspSerValTrpGly----- 329
Db 37602 AAAGATACGGTCTTAGGAAGGTTAGCTCTTCTTCAAGAAAGAGACAAACTCATTA 37661
QY 330 -----LeuSerCysAspThrValArg 336
Db 37662 GAAATGCACAGGCTTTATTGGAAATACAGTCTCAGCAGCTCTGTGCAAAAGTCTAA 37721
QY 337 -----GluGlnLeu 339
Db 37722 CTAGCTCTAGAGGCTTACTGAAGACAAGAAAGTTAGTGAAGAAATGAATCTTTG 37781
QY 340 ThrAsnSer-----IleArgLys 345
Db 37782 AAATCTTTAAGATTGCAGAAAGTACTAGTGGCAGAGAAACACAGAGAGCTACAAAA 37841
QY 346 GlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHis 365
Db 37842 GAGTATGAATTTCTTCGAGTCCCTATGAGATGTTAGTAATGAAGCAGAAAGGATTG 37901
QY 366 Ser-----GluGlyLeuAsnGluGlu-----AspValIleSerArg 377
Db 37902 CATGTGGTGAAGCTGTGAGCAAGAGAAACAAAGAACTGTATGGCAAGTTAAGAGACA 37961
QY 378 GlnAspHisGluGlnGluThrGluLys-----LeuGluLeuGluIleGluArg 393
Db 37962 GAGGCAAAACAAGAGAGACAGAAAAGCAGTTGCGAGGAGCTGAGCAAGAAATGGAGAA 38021
QY 394 CysLysGluMetIleLys-----AlaGlnGlnGlnLeuLeuGlnGln 408
Db 38022 ATGAAGAAAGAAATGAGAAAGTTTGTCTAAATCTAAACAGCAGAAATCTTAGAGCTG 38081
QY 409 -----LeuAlaThrThrCys 413
Db 38082 GAAGAGAATACCGGCTTAGGGCAGAGGTGCACCCCTGCAGGAGATACAGCTAAGAGT 38141
QY 414 AspAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLysGluArgLeu 433
Db 38142 ATGGAA-----ACACTCTTTCTTCCCAATGCCAGCATGAAGAGAACTTGAAGGTC 38195
QY 434 LysGluGluTrpThrLeuPheLysGluGlnLysLysAsnPheGluArgGluArgSer 453
Db 38196 AAAATGGAGTATGAACCCCTTTCTAAGAAAGTTTCAAGTCTTTAATGCTGTGAGAAAG 38255
QY 454 PheThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArgAla 473
Db 38256 CTAAGTGAAGAGGTTCAAGATTTAAAGCATCAGATAGAAGTAATGATATCTATCAACAG 38315
QY 474 SerTrp---ValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGlu 492
```

```
Db 38316 AACCTAGAGCCACCAGAGAAACATGATAACCAACGAATGTCTCACTGAAGAGGAAACACAG 38375
QY 493 AsnValLysLeuPheSerAlaPheSerGlySerAspProAspAsnLeuIleValHis 512
Db 38376 TCTATACCA-----GGTGAGACTGAAGACAGACTCTCTGAGTATGAGC 38420
QY 513 SerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCysThrSer 532
Db 38421 ACAAGA-----CCTACATGTTTCAGAA 38441
QY 533 LysLeuThrLysSerLeuProAlaSerProSerThrSer---AspPheArgGlnThrHis 551
Db 38442 TCGGTTCCATCAGCCGAGAGAGTGCACCTGCTGTAAGTAAGGATTTC----- 38489
QY 552 SerCysValSerGluHisSerIleSer----- 561
Db 38490 -----AGTCACATGATGAATTAATACTACACAGAGATTGATCAGCTCAA 38540
QY 562 -----ValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSer 579
Db 38541 GAAAGAATTTGCTGGATTAGAGGAGGAGAGCAAGAAAAACAAGGAATTTAGCCAGACTTTA 38600
QY 580 ThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGly 595
Db 38601 GAAATGAGAAAAATACCTTACTGAGTCAGATATCAACAAAGGATGGT 38648

RESULT 8
US-09-949-016-2009
; Sequence 2009, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2009
; LENGTH: 6276
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2009

Alignment Scores:
Pred. No.: 1,72e-12 Length: 6276
Score: 220.50 Matches: 157
Percent Similarity: 36.2% Conservative: 118
Best Local Similarity: 20.7% Mismatches: 233
Query Match: 7.0% Indels: 251
DB: 3 Gaps: 36

US-10-644-084-2 (1-615) x US-09-949-016-2009 (1-6276)
QY 16 AsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSer 35
Db 3033 AGTAAGCAAGTTAAAGAATATCGAATTCGAAGCAACCAATAGGAAATTTCTTCACCTC 3092
QY 36 -----GlnGlnValLeuCysSerSerValProLeuSer 46
Db 3093 AGTATTGAAGCTATTTTGGGAAGAATGTAAAGAGATTGTGAAGGCCTCT-----TCC 3143
QY 47 LysAsnValHisGlyValPheGlyValPheCysThrGlyGluAsnIleGluGlnSerIle 66
Db 3144 AAAAAAAGTCATCAGATT-----GAGGAACCTGGAACACAAATTT 3182
```


; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1028
 ; LENGTH: 6284
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-949-016-1028

Alignment Scores:
 Pred. No.: 1,72e-12 Length: 6284
 Score: 220.50 Matches: 157
 Percent Similarity: 36.2% Conservative: 118
 Best Local Similarity: 22.7% Mismatches: 233
 Query Match: 7.0% Indels: 251
 DB: 3 Gaps: 36

US-10-644-084-2 (1-615) x US-09-949-016-1028 (1-6284)

QY 16 AsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSer 35
 DB 3033 AGTAAGCAAGTTAAAGAAATATCGAATTCAAGAACCCCAATPAGGAAATAATCTTTCCACTCT 3092
 QY 36 -----GlnGlnValLeuCysSerSerValProLeuSer 46
 DB 3093 AGTATTGAAGCTATTGGGAAGAAATCTAAAGAGATTGTGAAGCCCTCT-----TCC 3143
 QY 47 LysAsnValHisGlyValPheGlyValPheCysThrGlyGluAsnIleGluGlnSerIle 66
 DB 3144 AAAAAAAGTCATCAGATT-----GAGGAAGTCGAACAACAAATTT 3182
 QY 67 SerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSerLeuTyrGluGlnSerLys 86
 DB 3183 GAAAAATTGCAAGGCAAGATAA-----GGCTATAAGGATGAAAAACAATAGACTAAAG 3236
 QY 87 SerLysGluAlaLysArgGluLeuAsnIleValAlaValLeuAsnCysMetAsnGluLeu 106
 DB 3237 GAGAGGAGCATAAAAACCCAGATGACCTACTAAAA-----GAAAAAGAAACT 3284
 QY 107 LeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSer-----ValGluThr 122
 DB 3285 CTTATACAGCAGCTGAAAGAGAAATTTGCAAGAAAAAATGTTACTCTTGATGTTCAATA 3344
 QY 123 GlnAsnLeuLysLeuGly-----SerAspMetAspHisLeuGlnSerCysTyr 138
 DB 3345 CAGCATGTAGTTGAAGGAAGAGAGCGCTTTTCAGAACTTACAAAGGTGTTACTGCTAT 3404
 QY 139 ---AlaLysLeuLysGluGln-----LeuGluThrSerArg----- 149
 DB 3405 AAGGCCAAAANTANAGGAACCTTGAAACAATTTTAGAGACTTCAGAAAGTTGAACGTAGTCAT 3464
 QY 150 -----ArgGluMetIleGlyLeuGlnGluArgAsp 159
 DB 3465 TCAGCCAAAGTTAGAACAGACATTTTGGAAAAAGGAATCTATC---ATCTTAAAGCTAGAA 3521
 QY 160 ArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAsp--- 178
 DB 3522 AGAAATTTGAAGGAATTTCAAGAACATCTTTCAGGATTCCTGTCAAAAAACACCAAGATTTA 3581
 QY 178 ----- 178
 DB 3582 AATGTAAAGGAACCTCAAGCTGAAAGAAATTCACAGATTAACAAATAATTTGCAAGAT 3641

QY 179 -----GluValGln 181
 DB 3642 ATGAAACATTACTTCAATTTAAAGAAAGAAAGAAACCAACCAAGCAAGCAAGCAAGAA 3701
 QY 182 LysLeu---GlnAsnIleIleAlaSerArgAlaThrGlnTyrAsn-----HisAspVal 198
 DB 3702 AAAATGAAAGAGGAACTCTCTGCAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3761
 QY 199 LysArgLysGluArgGluTyrAsnLysLeuLysGluArgLysGluHisGlnLeuValMetAsn 218
 DB 3762 CAGAGGAGGAGAGAGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3815
 QY 219 LysLysAspLysAsnIleAlaMetAspValLeuLeuAsnTyrValGlyArgAlaAspGlyLys 238
 DB 3816 AAGCAGATTAAAGCAAGTACAGAAAGAGGTA----- 3845
 QY 239 ArgGlySerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIle 258
 DB 3846 -----TCTGTAATGCGTGTGATGAGGATAAATTTACTGAGGATT 3881
 QY 259 LeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLys 278
 DB 3882 AAAATTATGAAGTGGAGAAAAAGAAACCAAGCTGTTCTCAGGAATTTAGATATGAAGCAG 3941
 QY 279 LysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLys 298
 DB 3942 CGAACCAATTCAGCAACTCAAGGAGCAGTTAAATAATCAGAAAGTGAAGCAAGCTATACAA 4001
 QY 299 ProArgGluArgAla-----GluAspGlyThr 307
 DB 4002 CAGTATGAGAGAGATGCAAAAGATCTAAATGTTAAAGAGAAATAATTTGAAGACATGCGA 4061
 QY 308 GlyThrValAlaIleSerAspIleGluAspSerGlyGluLeuSerArgAspSerVal 327
 DB 4062 ATGACACTAGAA-----GACAGGAAACAACTCAGTGAACAGGATCAAGTG 4109
 QY 328 TrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTyr 347
 DB 4110 CTTGAGGCTAAATTAGAGGAAGTT---GAAAGCTGGCCACAGAAATTTGAAAAA---TGG 4163
 QY 348 ArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGlu 367
 DB 4164 AAGGAAAAATGCAATGATTTCGAAACCAAAACCAATCAAGGCTCAATTAAGAACAATGAG 4223
 QY 368 GlyLeuAsnGluGluAspValIleSerArg-----GlnAspHisGluGlnGlu 383
 DB 4224 -----AACACACAGATGCTGCTGGAAGCTCACTAATCTTCAAGATGAGTTACAGGAG 4277
 QY 384 ThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAlaGlnGln 403
 DB 4278 TCTGAACAG---AAATATAATGCTGATAGA---AAGAAATGCTTAGAGAAAAAATGATG 4331
 QY 404 LeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspThrThrSerLeuLeuArgAsp 423
 DB 4332 CTTATCACTCAA----- 4343
 QY 424 CysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTyrThrLeuPheLysGluGln 443
 DB 4344 -----GCCAAAGAGCAGAGAAATATACGAAT-----AAGAGATG 4379
 QY 444 LysLysAsnPheGluArgGluArgSerPhe----- 454
 DB 4380 AAAAAATATGCTGAGGACAGGAGCGTTTTTTTAAAGCAACAGAAATGAATGGAATACTCTG 4439
 QY 455 ThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArg----- 472
 DB 4440 ACAGCCAGCTGACAGAGAAAGATAGTGACCTTCAAAAGTCCGCGAGAGAAACGATCAA 4499
 QY 473 -----AlaSerTyrValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGln 489
 DB 4500 CTGTTGCGACGCTTTTGAATATACAGTAAAGCACTGATATCCAGTAAATGTACAGAAAGAT 4559

```
Qy 490 Asn-----SerGluAsnValLysLeuPheSerAlaPhe 500
    |||
    |||
    |||
Db 4560 AATGAATTGAACAATAAAAGGATCATATCAGAGACTTCTAAATAGAAACACAAATC 4619
    |||
    |||
    |||
Qy 501 Ser-----GlySerSerAspProAsp----- 507
    |||
    |||
    |||
Db 4620 ATGGATATCAAGCCCAACGATTAGTTTCAGCAGATCCTGCACAACTTCAAACTGAACCT 4679
    |||
    |||
    |||
Qy 507 ----- 507

Db 4680 CTATCGACAAGTTTTTGAAATTTCCAGAAATAAAATAGAGATGGATCTGTAGTCCTTGAC 4739

Qy 508 -----AsnLeuile 510
    |||
    |||
    |||
Db 4740 TCTTGTGAAGTCTCAACAGAAATGATCAAGCACTCGATTCCAAAACCTGAGTTAGAG 4799

Qy 511 ValHisSerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCys 530
    |||
    |||
    |||
Db 4800 ATTCAATTTACACCTTTACAGCCAAACAAAATGGCAGTGAACAC-----CCTGGTTGT 4853

Qy 531 ThrSerLysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArgGlnThr 550
    |||
    |||
    |||
Db 4854 ACCACACAGTGACAGTTGAGATTCCTCCAGGCT-----CGGAAGAGG 4895

Qy 551 HisSerCysValSerGluHisSerSerile-----SerValLeuAsnIleThrPro 567
    |||
    |||
    |||
Db 4896 AAGAGTAATGAATGAGGAGGAGCTTGGTGAATGGAATGAAGAAATGACGTACACCC 4955

Qy 568 GluGluSerLysPro-----SerGluValAlaArgGluSerThrAspGlnLysTrp 584
    |||
    |||
    |||
Db 4956 AGAAGTAATTTGAAATTTCTTATTTTCAGATAGATAAAATTTCTTGTCAAAAAGGAACAA 5015

Qy 595 SerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCysSerSerAla 603
    |||
    |||
    |||
Db 5016 AAGGTGGCCATACGTCATCTAAGAAA---ACATATTCTTTACGGAGTCAGGCA 5069
```

RESULT 10

```
US-09-643-597-117
; Sequence 117, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 117
; LENGTH: 6921
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-643-597-117
```

```
Alignment Scores:
Pred. No.: 4,86e-12 Length: 6921
Score: 217.00 Matches: 144
Percent Similarity: 38.4% Conservative: 140
Best Local Similarity: 19.5% Mismatches: 268
Query Match: 6.9% Indels: 187
DB: 3 Gaps: 29
```

```
US-10-644-084-2 (1-615) x US-09-643-597-117 (1-6921)
Qy 1 MetGlyAspTyrMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
    |||
    |||
    |||
Db 1265 TTAGATGATTGGATC-----CAGCAGGTTTGAAACTACTCTCAGAGAAAGATTCCAGGAA 1315
    |||
    |||
    |||
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
    |||
    |||
    |||
Db 1316 AATCAGCCTGAAATAGTAAACCTAGCCACACAGTTTGAATCAACAGAGATGCTGGTG 1375

Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
    |||
    |||
    |||
Db 1376 TCCGAAATAGAAATGAAACAGAGCAAAATGACGAG-----TGTCAAAATAT 1423

Qy 61 AsnIleGluGlnSerIleSerLysLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
    |||
    |||
    |||
Db 1424 GCAGAAACAGTACTCAGCTACAGTGAAGCATATGAATTTACAAACATGACCTACCGGCC 1483

Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArg----- 92
    |||
    |||
    |||
Db 1484 ATGCTAGATTCAACAACAAATCTCCAGTGAACCGCGAAGAAATGCAGAGTTTCAGCAGAT 1543

Qy 93 -----GluLeuAsnIleValAlaValLeuAsnCysMet 103
    |||
    |||
    |||
Db 1544 CTCATTATTCAAGAGTTTCATGACCTAAGGACTCGATATCTGCTGGTGTCTCTCATG 1603

Qy 104 AsnGluLeuValLeuGlnArgLysAsnLeuAlaGlnGluSerValGluThrGln 123
    |||
    |||
    |||
Db 1604 ACACAAATATATAATTTGCTGGTGATTCATTGAAGAGGCTGGAAGAGGAGGAGATTAAA 1663

Qy 124 AsnLeuLysLeuGlySerAsp-MetAspHisLeuGlnSerCysTyrAlaLysLeuLysGlu 143
    |||
    |||
    |||
Db 1664 AGGTGTAAGGAGACTTCTGAACATGGGCATATTTCAGATCTGCTTCACGCTCAGAAGCA 1723

Qy 143 uGln--LeuGluThrSerArg-----ArgGluMetIle 153
    |||
    |||
    |||
Db 1724 ACAGTGTCTGAGAAATACGAAACTTACAGGAAGATAAGTGAGTTGGAAAGAAATGGTAGCT 1783

Qy 154 GlyLeuGlnGluArgAspArgGlnGluCysLysAsnArgSerLeuHisGlnLeuLeu 173
    |||
    |||
    |||
Db 1784 GAACATAAGAAACAAAGTCCCGAGTAGAGAGAAATCTAGAGATATCTCTCTGCAAGATAAGG 1843

Qy 174 LysAsnGluLysAspGluValGlnLys-----LeuGlnAsnIleIle 187
    |||
    |||
    |||
Db 1844 GAAATGAATTTGACAAAGCAGCAGACAGAAATCTAGAGATATCTCTCTGCAAGATAAGG 1903

Qy 188 Ala---SerArgAlaThrGlnTyrAsnHisAsp-----ValLysArgLysGlu 202
    |||
    |||
    |||
Db 1904 GCTGAAAGTGAAGCCCAAGCAGTACCCGAGGGAACCTTGAACCACTTTGTGAGAGAGAAAGAA 1963

Qy 203 ArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLys 222
    |||
    |||
    |||
Db 1964 GCCCTGAAAGAGAACTG---GAGCGGTGAGGAGCTCACCATAGAGCCGCGAGCTAAA 2020

Qy 223 AsnIleAlaMetAsp-----ValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
    |||
    |||
    |||
Db 2021 AGAGCTGCGGTGGAAGAGAACCTCTGCAATTTTCGCAATCAGTTGGAGGAAAAACACCTTT 2080

Qy 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
    |||
    |||
    |||
Db 2081 ACCAGACGAACACTGGAAGATCATCTTAAAGAAAAGAT-----TTAAGTCTC 2128

Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGlu----- 273
    |||
    |||
    |||
Db 2129 AATGATTTGGAG---CAACAAAAAATAAATAATATGAGAAATTAAGAGAAAGAGAGAC 2185

Qy 274 ---AsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIle----- 289
    |||
    |||
    |||
Db 2186 AATCAGGAAGAACTCTTGAAGCTGATAAGAGCAGATGGAAGAAAGACCTTTGCATTTTCA 2245

Qy 289 ----- 289
Db 2246 CAGGTAGCAGAGAAACAGTTGAAAGAAAGCAGAAAAATTTGAATTGGAAAGCAAGAGAAA 2305
```

```
QY 289 ----- 289
Db 2306 ATAACTGAAATTCAGTATACATAGAGAGAAATGCCATTGCCAGTGTGTCGGATCACACAG 2365
QY 290 -----SerLeuLeuSerProGlnLysLysProArgGluArgAlaGlu 304
Db 2366 GCTACATCATGCAGGCGCAGTAAGCGGTCTCCAGCAGACATGACACAGCAGAGAGAGAGAA 2425
QY 305 Asp-----GlyThrGlyThrValAlaIleSerAspIleGluAspAspSer 319
Db 2426 GAACTCAAAACAGCAGGTAGTAACTAAACAGTCCCAATAGAAAGGCTGAACAGACATG 2485
QY 320 GlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeu 339
Db 2486 AGAGAGCTGACATATGAA-----CTTAATGCCCTCCAGCTTGAAAAA 2527
QY 340 ThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsn 359
Db 2528 AGCTCATCTGAGGAAAGGCT---CGTTTGCTAAAGATAAACTAGATGAACAAATAAT 2584
QY 360 -----GlnAlaSerLysValHisSer 366
Db 2585 ACACTCAGATGCTTAAGTTGGAGCTGGAAAGGATCAGGCGGAGAAAGGTTATCT 2644
QY 367 GluGlyLeuAsnGlu-----GluAspVal 374
Db 2645 CAACAACTCAGAGAGCTTGGTAGGCAATTGAATCAAAACACACAGGTAAAGCTGAAGAGCC 2704
QY 375 IleSer-----ArgGlnAspHisGluGlnGluThrGlyLys 386
Db 2705 ATGCAGAGAGCTAGTATCTCAAGAAAATAAAGCGCAATTATCAGTTAGTAATATCT 2764
QY 387 LeuGluLeuGlu-----IleGluArgCysLysGlu 396
Db 2765 CTTAATCATGAAAAGGGAAACTACAAAGAGAAGTAGACAGAAATCAACAGGGCACATGCT 2824
QY 397 MetIleLysAlaGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAsp 416
Db 2825 GTAGCTGAGAGAATAATTCAGCATTTAAATTCACAAATTCATCTTTTCAGATGAGAA 2884
QY 417 ThrThrSerLeuArgAspCysTyrLeuLeuGluLysGluArgLysGluGlu 436
Db 2885 GAATTAGAAAGACTACAATCTGCCAG-----AGNAATCAGATCATCTAAAGAACAA 2938
QY 437 TrpThrLeuPheLysGluGln---LysLysAsnPheGluArgGluArgSerPheThr 455
Db 2939 TTTGAGAAAAGCCATGAGCAGTTGCTTCAAAATATCAAAAGCTGAAAAGAGAAAT---AAT 2995
QY 456 GluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrp 475
Db 2996 GATAAAATCCAAAGGCTCAATGAAGAAATTTGGAGAA---AGTAATGAGTGTGCAGAGATG 3052
QY 476 ValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLys 495
Db 3053 CTAATAACAAAGAGTAGAGAGCTTACT-----AGCCAGAAATAAGAACCAAAATTA 3103
QY 496 LeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgPro 515
Db 3104 ATGATGCAGAGAAATTCAGGCAGAAATCA-----GAGAAATATAGTTTAGAGAAACAAACT 3157
QY 516 ArgGlnLysLysLeuHisSerVal-----AlaAsnGlyValProAlaCysThrSer 532
Db 3158 ATCCAGCAAAAGATGTGAAGCACTGAAATTCAGGCAGATGGTTTAAAGATCAGCTACGC 3217
QY 533 LysLeuThrLysSerLeuProAlaSerProSerThr---SerAspPheArgGlnThrHis 551
Db 3218 ACCCAAAATGACATCTGCATTAACACAGAAAACAGACGAGATTTTCAAAAGAAAAT 3277
QY 552 SerCysValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGluSerLys 571
Db 3278 AAATGCTTAGAGAA-----GACCTGGCGAAAGATCAAAATTTG 3316
```

```
QY 572 ProSerGluValAlaArgGluSerThrAspGlnLysThrSerValGlnSer 588
Db 3317 GTAAGTGAATTTAAGCAAAAGTGTGACCAACAGAACATTATCATCCAGAAT 3367
RESULT 11
US-09-480-884A-117
; Sequence 117, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Ligu
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; FILE REFERENCE: 210121.455C6
; CURRENT APPLICATION NUMBER: US/09/480,884A
; CURRENT FILING DATE: 2001-08-27
; NUMBER OF SEQ ID NOS: 330
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 117
; LENGTH: 6921
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-480-884A-117
Alignment Scores:
Pred. No.: 4-86e-12 Length: 6921
Score: 217.00 Matches: 144
Percent Similarity: 38.4% Conservative: 140
Best Local Similarity: 19.5% Mismatches: 268
Query Match: 6.9% Indels: 187
DB: 3 Gaps: 29
US-10-644-084-2 (1-615) x US-09-480-884A-117 (1-6921)
```

```
QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
Db 1265 TTAGATGATTGGATC-----CAGCAGGTTGAAACTACTCAGAGAAAAGATTTCAGGAA 1315
QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db 1316 AATCAGCTGAAATATGATAAAACCTCAGCCACACAGATTGAAATCAACAGAGATGCTGGTG 1375
QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 1376 TCCGAAATAGAAATGAAACAGAGACAAAATGGACGAG-----TGTCAAAATAT 1423
QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
Db 1424 GCAGAACAGTACTCAGCTACAGTGAAGGACTATGAATTACAACAAATGACCTACCGGCC 1483
QY 81 LeuTyrGluGluSerLysSerLysGluAlaLysArg----- 92
Db 1484 ATGGTAGATTCAACAAACAAATCTCCAGTGAACCCCGAAGAAATGCAGAGTTCAGCAGAT 1543
QY 93 -----GluLeuAsnIleValAlaValLeuAsnCysMet 103
Db 1544 CTCATTATTCAGAGTTCATGACCTAAGGACTCGATATCTGCTGCTGCTCTCATG 1603
QY 104 AsnGluLeuLeuValLeuGlnArgLysAsnLeuAlaGlnGlnSerValGlnThrGln 123
Db 1604 ACACAAATATATAAATTTGCTGTTGATTCAITTAAGAGGCTGGAAGAGGAGGAGAGATTA 1663
QY 124 AsnLeuLysLeuGlySerAsp-MetAspHisLeuGlnSerCysTyrAlaLysLeuLysGln 143
Db 1664 AGGTGAAGAGAGACTTCTGAACATGGGCAATATTCAGATCTGCTTCAAGCGTCAGAGGCA 1723
QY 143 uGln--LeuGluThrSerArg-----ArgGluMetIle 153
Db 1724 ACAGTCTTGAGATAGCAAACTTACAGGAAAGATAGTAGTTCGAAAGAAATGTTAGCT 1783
```


Qy	1	MetGlyAspTrrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer	20
Db	1265	TTAGATTTGGATC-----CAGCAGGTTGAACTACTCAGAGAAAGATTTCAGAA	1315
Qy	21	GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys	40
Db	1316	AATCAGCCTGAAATAGTAAACCTTAGCCACACAGITTGAAATCAACAGAGATGCTGGT	1375
Qy	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu	60
Db	1376	TCCGAATAGAAATGAACACAGACAGAAATGACGAG-----TGTCAAANAATAT	1423
Qy	61	AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer	80
Db	1424	GCAGAACAGTACTCAGCTACAGTGAAGGACTATGAATTACAAACAATGACTACCGGCC	1483
Qy	81	LeuTyrGluGluSerLysSerLysGluAlaLysArg-----	92
Db	1484	ATGTTAGATTCCAAACAAAATCTCCAGTGAACGCCGAAAGATGCGAGATTTCAGCAGAT	1543
Qy	93	-----GluLeuAsnIleValAlaValLeuAsnCysMet	103
Db	1544	CTCATTATTCAAGATTTCATGGACCTTAAGGACTCGATATATCTGCCCTGGTCACTCTCATG	1603
Qy	104	AsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerValGluThrGln	123
Db	1604	ACACAATATATTAATTTGCTGGTGATTTCATTGAAGAGGCTGGAAGAGGAGGAGATTAAA	1663
Qy	124	AsnLeuLysLeuGlySerAsp-MetAspHisLeuGlnSerCysTyrTrpAlaLysLeuLysGI	143
Db	1664	AGGTGAAGGAGACTTCTGAACATGGGGCATATTTCAGATCTGCTTCAGCGTCAGAAGGCA	1723
Qy	143	uGln--LeuGluThrSerArg-----ArgGluMetIle	153
Db	1724	ACAGTCTTGAGATAGCAAACTTACAGGAAGATAAGTCAAGTGGAAAGATGGTAGCT	1783
Qy	154	GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu	173
Db	1784	GAACATAAGAAACAAAAGTCCGAGTAGAGGAAGAACTTCCGAAGGTCCAGGAGGCTGCA	1843
Qy	174	LysAsnGluLysAspGluValGlnLys-----LeuGlnAsnIleIle	187
Db	1844	GAATAATGAATTGAGAAAGCAGACAGAGAAATGTAGAAGATATCTCTTCGACAGAGATAAGG	1903
Qy	188	Ala---SerArgAlaThrGlnTyrAsnHisAsp-----ValLysArgLysGlu	202
Db	1904	GCTGAAAGTGNAGCCAAGCAGTACCGCAGGGAATCTGAAACCATTTGTGAGAGAGAAGGAA	1963
Qy	203	ArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLys	222
Db	1964	GCCGCTGAAAGAGAACTG---GAGCGGGTGGAGGAGCTCACCATGAGGCGGAGGCTTAA	2020
Qy	223	AsnIleAlaMetAsp-----ValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly	240
Db	2021	AGAGCTGCCGTGGAAGAGAACTCTCTGAAATTTTCGCAATCAGTTGGAGGAAAAACACCTTT	2080
Qy	241	SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu	260
Db	2081	ACCAGACGAACCTGGAGAGATCATCTTTAAAGAAAAAGAT-----TTAAGTCTC	2128
Qy	261	AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGlu-----	273
Db	2129	AATGATTTGGAG---CAACAAAAAATAAATTAATGGAAGAAATTAAGAGAAAGAGAGAC	2185
Qy	274	---AsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIle-----	289
Db	2186	AATGAGGAAGAACTCTTTGAAGCTGTATAAGCAGATGAAAAAGACCTTGATTCATTCAGAAA	2245
Qy	289	-----	289
Db	2246	CAGGTACGAGAAACAGTTGAAAGAAAAACGAGAAATTAATTTGAGAACAGAGAAAA	2305
Qy	289	-----	289

2306	DB	ATAACTGA	AAATTCAGTATACATGATGAGAGAAAATGCAATGTCAGTGTGTCGATCCACACAG	2363
290	QY	-----	SerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGlu	304
2366	DB	GCTACATCATGCAGCGCAGTAACGGGTCTCCAGCAAGAACATGACACAGCAGAAAGCAGAA	2425	
305	QY	Asp-----	GlyThrGlyThrValAlaIleSerAspIleGluAspAspSer	319
2426	DB	GAACCTCAACACAGCAGGTAGATGAACCTAACAGCTGCCAATAGAAGGCTGAACACAGACATG	2485	
320	QY	GlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeu	339	
2486	DB	AGAGAGCTGACATATGAA-----	CTTAATGCCCTCCAGCTTTGAAAAA	2527
340	QY	ThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsn	359	
2528	DB	ACGTCACTCTGAGGAAAAGGCT--CGTTTCTTAAAGATATAACTAGATGAACAAATAAT	2584	
360	QY	-----	GlnAlaSerLysValHisSer	366
2585	DB	ACACTCAGATGCCCTTAAGTTGGAGCTGGAAAGGAAGGATCAGCGGGAAGAAGGGTATTCT	2644	
367	QY	GluGlyLeuAsnGlu-----	-----GluAspVal	374
2645	DB	CAACAACTCAGAGAGCTTGTGTAGCAATTCGATCAAAACCCACAGAGTAAAGCTGAAGAACCC	2704	
375	QY	IleSer-----	ArgGlnAspHisGluGlnGluThrGluLys	386
2705	DB	ATGCAAGAAGCTAGTGATCTCAAGAAATAAAGCGCAATTATCAGTTAGAAATTAGAATCT	2764	
387	QY	LeuGlnLeuGlu-----	IleGluArgCysLysGlu	396
2765	DB	CTTAATCATGAAAAAGGAAACTCAAAAGAGAAAGTAGACAGAAATCAACAGGGCAGCATGCT	2824	
397	QY	MetIleLysAlaGlnGlnLeuLeuGlnGlnGlnLeuAlaThrCysAspAspAsp	416	
2825	DB	GTAGCTGAGAAGAAATATTCAGCATTTAAATTCACAAATTCATCTCTTTCCAGATGAGAAA	2884	
417	QY	ThrThrSerLeuLeuArgAspCysTyrrLeuLeuGluGluGluArgLysGluGlu	436	
2885	DB	GAATTAGAAAGACTTACAAATCTGCCAG-----	AGAAATTCAGATCATCTATAAAGAACAA	2938
437	QY	TrpThrLeuPheLysGluGln-----	LysLysAsnPheGluArgGluArgArgSerPheThr	455
2939	DB	TTTGAGAAAAAGCCATGACGAGTGTCTTCAAAATATCAAAAGCTCAAAAGAAAAAAT--	AAT	2995
456	QY	GluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrp	475	
2996	DB	GATAAAATCCAAAGGCTCAATGAAGAATTGGAGAAA--	AGTAATGAGTGTGCAGAGATG	3052
476	QY	ValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLys	495	
3053	DB	CTAAAAAATAAAGTAGAGAGCTTACT-----	AGGCAGAAATAATGAACCAAAATTA	3103
496	QY	LeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgPro	515	
3104	DB	ATGATGCAGAGAATTCAGCGAGAATCA-----	GAGAAATATAGTTTTAGAGAAACAAACT	3157
516	QY	ArgGlnLysLysLeuHisSerVal-----	AlaAsnGlyValProAlaCysThrSer	532
3158	DB	ATCCAGCAAGAGATGTGAAGCACTGAAAAATTCAGCGAGATGGTTTTAAAGATCAGCTACGC	3217	
533	QY	LysLeuThrLysSerLeuProAlaSerProSerThr-----	SerAspPheArgGlnThrHis	551
3218	DB	AGCACAATGAACACTTGCATAAACACAGACAAAAACAGACGAGGATTTTCAAGAAAAAAT	3277	
552	QY	SerCysValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGluSerLys	571	
3278	DB	AAATGCTAGAGAA-----	GACTGCCGAAAAAGTCAAAATTTG	3316
572	QY	ProSerGluValAlaArgGluSerThrAspGlnLysTrpSerValGlnSer	588	

```

Db      3317 GTAAGTGAATTTAAGCAAAAGTGTGACCAACAGAAACATTATCATCCAGAAT 3367
RESULT 13
US-09-606-421B-117
; Sequence 117, Application US/09606421B
; Patent No. 6531315
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF LUNG CANCER
; FILE REFERENCE: 210121.455C9
; CURRENT APPLICATION NUMBER: US/09/606,421B
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 117
; LENGTH: 6921
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-606-421B-117

Alignment Scores:
Pred. No.:      4,866-12      Length:      6921
Score:          217.00      Matches:      144
Percent Similarity: 38.4%      Conservative: 140
Best Local Similarity: 19.5%      Mismatches:   268
Query Match:      6.9%      Indels:       187
DB:              3          Gaps:         29

US-10-644-084-2 (1-615) x US-09-606-421B-117 (1-6921)
QY      1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
      :|||:|||||:|
Db      1265 TTAGATGATGGATC-----CAGCAGGTGAAACTACTACAGAAAGATTCAGAA 1315
QY      21 GlnTyrThrSerGluThrLysMetSerProSerLeuTyrSerGlnGlnValLeuCys 40
      :|||:|||||:|
Db      1316 AATCAGCTGAAATAGTAAACCTTAGCCACACAGTTGAATCAACAGAGATGCTGTG 1375
QY      41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
      :|||:|||||:|
Db      1376 TCCGAAATAGAAATGAACAGAGCAAAATGGACGAG-----TGTCAAAAATAT 1423
QY      61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
      :|||:|||||:|
Db      1424 GCAGAACAGTACTACGACTAGCTAGGAGCTATGAAATTACAAACAATGACCTACCGGCC 1483
QY      81 LeuTyrGluGluSerLysSerLysGluAlaLysArg----- 92
      :|||:|||||:|
Db      1484 ATGGTAGATTCAACAAAAATCTCCAGTGAACCCGCAAGAAATCCAGATTCCAGCAT 1543
QY      93 -----GluLeuAsnIleValAlaValLeuAsnCysMet 103
      :|||:|||||:|
Db      1544 CTCATTATTCAAGAGTTTCATGGACCTAAGGACTCGATATACTGCGCTGCTCCTCATG 1603
QY      104 AsnGluLeuValLeuGlnArgLysAsnLeuAlaGlnGluSerValGluThrGln 123
      :|||:|||||:|
Db      1604 ACACAATATATTAATTTGCTGGTGATTCATTGAAGAGCGCTGGAAGAGGAGAGATATA 1663
QY      124 AsnLeuValLeuGlySerAsp-MetAspHisLeuGlnSerCysTyrAlaLysLeuLysGln 143
      :|||:|||||:|
Db      1664 AGGTGTAAGGAGACTTCTGAACATGGGCATATTCAGATCTGCTTCACGGCTCAGAGGCA 1723
QY      143 uGln--LeuGluThrSerArg-----ArgGluMetIle 153
      :|||:|||||:|

```

```

Db      1724 ACAGTGTGTTGAGAAATAGCAAACTTACAGGAAGAGATAAGTAGTTGGAAAGAAATGCTAGCT 1783
QY      154 GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu 173
      :|||:|||||:|
Db      1784 GAACCTAAAGAAACAAAGAGTCCGGAGTAGAGGAAGAACTTCCGAAGGCTCAGGAGGCTGCA 1843
QY      174 LysAsnGluLysAspGluValGlnLys-----LeuGlnAsnIleIle 187
      :|||:|||||:|
Db      1844 GAAATGAATTGAGAAACAGCAGAGAAATGTAGAAGATATCTCTCTGCAGAAAGATAAGG 1903
QY      188 Ala---SerArgAlaThrGlnTyrAsnHisAsp-----ValLysArgLysGlu 202
      :|||:|||||:|
Db      1904 GCTGAAAGTGAAGCCAGCAGTACCGCAGGAGAACTTGAAACATTGTGAGAGAGAGGAA 1963
QY      203 ArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLys 222
      :|||:|||||:|
Db      1964 CCCGCTGAAAGAGAACTG---GAGCGGTGAGGAGCTCACCATAGAGCCGAGGCTAAA 2020
QY      223 AsnIleAlaMetAsp-----ValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
      :|||:|||||:|
Db      2021 AGAGCTGCCGTGGAAGAGAACTCTCTGAATTTTCGCAATCAGTTGGAGGAGAAACACCTTT 2080
QY      241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
      :|||:|||||:|
Db      2081 ACCAGACGAACTGGAAGATCATCTTTAAAGAAAGAAAT-----TTAAGTCTC 2128
QY      261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGlu----- 273
      :|||:|||||:|
Db      2129 AATGATTTGGAG---CAACAAAAAATAATTAATGGAAGAAATTAAGAAGAAAGAGAGAC 2185
QY      274 ---AsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIle----- 289
      :|||:|||||:|
Db      2186 AATGAGGAAGAACTCTTTGAGCTGATAAGCAGATGGAAGAAAGACCTTTCGATTTTCAGAAA 2245
QY      289 ----- 289
      :|||:|||||:|
Db      2246 CAGGTAGCAGAGAAACACAGTTGAAAGAAAGCAGAAATTTGAATTGGAGAGCAAGAGAAAA 2305
QY      289 ----- 289
      :|||:|||||:|
Db      2306 ATAAGTGAATTCAGTATACATGTAGAGAAATGCAATTCGATTCGCGATGCACACAG 2365
QY      290 -----SerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGlu 304
      :|||:|||||:|
Db      2366 GCTACATCATCGCGGCGAGTAAACGGGTCTCCAGCAAGAACATGACACAGCAGAAAGCAGAA 2425
QY      305 Asp-----GlyThrGlyThrValAlaIleSerAspIleGluAspAspSer 319
      :|||:|||||:|
Db      2426 GAACCTCAACACAGCAGGTAGATGAACTAACAGCTGCCAATAGAAAAGGCTGGAACAAGCATG 2485
QY      320 GlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeu 339
      :|||:|||||:|
Db      2486 AGAGAGCTGACATATGAA-----CTTAATGCCCTCCAGCTTGAAGAAA 2527
QY      340 ThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsn 359
      :|||:|||||:|
Db      2528 ACGTCATCTGAGGAAAGGCT---CGTTTGCTAAAAGATAAACTAGATGAAACAAATAAT 2584
QY      360 -----GlnAlaSerLysValHisSer 366
      :|||:|||||:|
Db      2585 ACACCTCAGATGCCTTAAGTTGGAGCTGGAAAGGAGGATCAGGCGGAGAAAGGTTATCT 2644
QY      367 GluGlyLeuAsnGlu-----GluAspVal 374
      :|||:|||||:|
Db      2645 CAACACTCAGAGAGCTTGGTAGCAATTGAATCAACACACAGGTAAAGCTGGAAGAGCC 2704
QY      375 IleSer-----ArgGlnAspHisGluGlnGluThrGluLys 386
      :|||:|||||:|
Db      2705 ATGCAAGAGCTAGTGATCTCAAGAAAAATAAAGCGCAATTTATCAGTTAGTAATTTAGAACT 2764
QY      387 LeuGluLeuGlu-----IleGluArgCysLysGlu 396
      :|||:|||||:|
Db      2765 CTTAATCATGAAAGGAGGAAACTACAAAGAGAGAGTAGACAGAAATCACAAGGCGCATGCT 2824

```

```
QY 397 MetIleLysAlaGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAsp 416
   :::::  :::::  |||||  :::::  :::::  :::::  :::::  :::::
Db 2825 GTAGCTGAGAGAATAATTGAGCATTTTAAATTCACAAATTCATTTTCGAGATGAGAAA 2884
QY 417 ThrThrSerLeuLeuArgAspCysThrLeuLeuGluGluGluGluArgLeuGluGlu 436
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 2885 GAATTAGAAAGACTACAAATCGCCAG-----AGAAATCATCATCTAAAGAACAA 2938
QY 437 TrpThrLeuPheLysGluGln---LysLysAsnPheGluArgGluArgSerPheThr 455
   :::::  :::::  |||||  :::::  :::::  :::::  :::::  :::::
Db 2939 TTTGAGAAAAGCCATGAGCAGTGTCTTCAAAATATCAAAAGCTGAAAAAGAAAAT--AAT 2995
QY 456 GluAlaAlaIleArgLeuGluGluArgLysAlaPheGluGluGluArgLysThr 475
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 2996 GATAAAATCCAAAGGCTCAATGAAGAAATTTGGAGAAA---AGTAATGAGTGCAGAGATG 3052
QY 476 ValLysGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLys 495
   :::::  |||||  :::::  :::::  |||||  :::::  |||||  :::::
Db 3053 CTAAACAAAGATGAGAGAGCTTACT-----AGCAGAAATTAATAAACCAAAATTA 3103
QY 496 LeuPheSerAlaPheSerGlySerAspProAspAsnLeuIleValHisSerArgPro 515
   :::::  |||||  :::::  :::::  :::::  :::::  :::::  :::::
Db 3104 ATGATGCAGAGAATTCAGCAGAAATCA-----GAGAAATATAGTTTACAGAGAAACAACT 3157
QY 516 ArgGlnLysLysLeuHisSerVal-----AlaAsnGlyValProAlaCysThrSer 532
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 3158 ATCCAGCAAAAGATGTGAAGCACTGAAAAATTCAGGCAGATGGTTTTAAAGATCAGCTACGC 3217
QY 533 LysLeuThrLysSerLeuProAlaSerProSerThr---SerAspPheArgGlnThrHis 551
   :::::  |||||  :::::  :::::  |||||  :::::  |||||  :::::
Db 3218 ACCACAAATGAACACTTGCATAAACAGACAGAAAAACAGACGAGATTTTCAAAGAAAAATTT 3277
QY 552 SerCysValSerGluHisSerIleSerValLeuAsnIleThrProGluGluSerLys 571
   :::::  |||||  :::::  :::::  :::::  :::::  :::::  :::::
Db 3278 AAATGCTAGAGAA-----GACCTGGCGAAAGTCAAAATTTG 3316
QY 572 ProSerGluValAlaArgGluSerThrAspGlnLysTrpSerValGlnSer 588
   |||||  :::::  :::::  :::::  |||||  :::::  :::::  :::::
Db 3317 GTAAGTGAATTTAAGCAAAAGTGTGACCAACAGAACATTTATCATCCAGAAAT 3367

RESULT 14
; Sequence 117, Application US/09221107
; Patent No. 6660838
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C2
; CURRENT APPLICATION NUMBER: US/09/221,107
; CURRENT FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 161
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 117
; LENGTH: 6921
; TYPE: DNA
; ORGANISM: Human
US-09-221-107-117

Alignment Scores:
Pred. No.: 4,86e-12 Length: 6921
Score: 217.00 Matches: 144
Percent Similarity: 38.4% Conservative: 140
Best Local Similarity: 19.5% Mismatches: 268
Query Match: 6.9% Indels: 167
DB: 3 Gaps: 29

US-10-644-084-2 (1-615) x US-09-221-107-117 (1-6921)

QY 1 MetGlyAspTyrMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
   :::::  |||||  :::::  :::::  :::::  :::::  :::::  :::::
Db 1265 TTATGATGATTGGATC-----CAGCAGGTTGAACACTACTCAGAGAAAGATTCAGGAA 1315
```

```
QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1316 AATGAGCCCTGAAATAATGATAAAACCCCTAGCCACACAGATTTGAATCAACAGAGATGCTGGTG 1375
QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
   |||||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1376 TCCGAATAGAAATGAACAGAGCAAAATGGACGAG-----TGTCAAATAATAT 1423
QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
   :::::  |||||  :::::  :::::  |||||  :::::  |||||  :::::
Db 1424 GCAGAACAGTACTCAGCTACAGTACAGTAAAGACTATGAATTAACAACAATGACCTACCGGGCC 1483
QY 81 LeuTyrGluGluSerLysSerLysGluAlaLysArg----- 92
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1484 ATGGTAGATTTCACAAACAAAAATCTCCAGTGAACCCGGAAGAAATCCAGAGATTCAGCAGAT 1543
QY 93 -----GluLeuAsnIleValAlaValLeuAsnCysMet 103
   |||||  :::::  :::::  :::::  |||||  :::::  |||||  :::::
Db 1544 CTCATTATTCAAGAGTTTCATGGACCTAGGACTCCATATACTGCTGCTGCTCACTCTCATG 1603
QY 104 AsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerValGluThrGln 123
   :::::  :::::  |||||  :::::  :::::  :::::  :::::  :::::
Db 1604 ACACAATATATTAAATTTGCTGTCATTTCATTGAAGAGGCTGGAAGAGGAGGAGATTAATA 1663
QY 124 AsnLeuLysLeuGlySerAsp-MetAspHisLeuGlnSerCysTyrAlaLysLeuLysGlu 143
   :::::  |||||  :::::  :::::  |||||  :::::  |||||  :::::
Db 1664 AGGTGAAGGAGACTTCTGAACATGGGCATATTCAGATCTCTCTCAGCGTCAGAGGCA 1723
QY 143 uGln--LeuGluThrSerArg-----ArgGluMetIle 153
   |||||  |||||  |||||  :::::  :::::  :::::  :::::
Db 1724 ACAGTCTTTGAGATAGCAAACTTACAGAAAGATAAGTGAGTTGGAAGAATGGTAGCT 1783
QY 154 GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu 173
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 1784 GAACATAAGAAACAAAGTCCCGAGTAGAGGAAGAATTCCTCGAAGGTCAGGAGGCTGCA 1843
QY 174 LysAsnGluLysAspGluValGlnLys-----LeuGlnAsnIleIle 187
   :::::  |||||  :::::  :::::  |||||  :::::  |||||  :::::
Db 1844 GAAATATGAATTTGAGAAAGCAGCAGAGAAATGTAGAAGATATCTCTCTGCAGAGATAAG 1903
QY 188 Ala---SerArgAlaThrGlnTyrAsnHisAsp-----ValLysArgLysGlu 202
   |||||  |||||  |||||  |||||  :::::  :::::  :::::  :::::
Db 1904 GCTGAAGTGAAGCCACAGCAGTACCGCAGGGAACCTTGAAACCATTTGTGAGAGAGAA 1963
QY 203 ArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLys 222
   :::::  :::::  |||||  :::::  :::::  :::::  :::::  :::::
Db 1964 GCGCGTGAAGAGAACTG--GAGCGGTGAGGCACTCACCATTAGAGSCCGAGGCTAAA 2020
QY 223 AsnIleAlaMetAsp-----ValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
   |||||  :::::  :::::  :::::  |||||  :::::  |||||  :::::
Db 2021 AGAGCTGCGTGGAAAGAGAACCTCTCTGAATTTTTCGCAATCAGTTGGAGAAAAACCTTT 2080
QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
   :::::  |||||  :::::  :::::  |||||  :::::  |||||  :::::
Db 2081 ACCAGACAAACACTGGAAGATCATCTTAAAGAAAAGAT-----TTAAGTCTC 2128
QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGlu----- 273
   |||||  |||||  |||||  :::::  :::::  :::::  :::::
Db 2129 AATGATTTGGAG---CAACAAAAATAAATTAATGGAAGAATTAAGAAGAAGAGAGAC 2185
QY 274 ---AsnAlaGluLeuLysLysValLeuGlnMetLysLysGluMetIle----- 289
   |||||  |||||  :::::  :::::  |||||  :::::  |||||  :::::
Db 2186 AATGAGGAAGAACTCTTGAAGCTGATAAAGCAGATGGAAAAAGACCTTGCATTTTCAGAAA 2245
QY 289 ----- 289
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 2246 CAGGTAGCAGAGAAACAGTTGAAAGAAAGCAGAAAAATTTGAATTCGAAGCAAGAGAAAA 2305
QY 289 ----- 289
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 2306 ATAACTGAAATTCAGTATACATGTAGAGAAAAATGCAATTCGCCAGTGTGTCGATCACACAG 2365
QY 290 -----SerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGlu 304
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
```


Db 1904 GCTGAAAGTGAAGCCAGCAGTAGTACCGCAGGAACTTGAACACCATTTGTGAGAGAGAGGAA 1963
Qy 203 ArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLys 222
Db 1964 GCCGCTGAAAGAGAACTG---GAGCGGCTGAGGCAGCTCACCATGAGGCGCGGCTAAA 2020
Qy 223 AsnIleAlaMetAsp-----ValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db 2021 AGAGCTGCCGTGAGAGAACTCCTGAAATTTTCGCNATCAGTTGGAGGAAAAACACCTTT 2080
Qy 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
Db 2081 ACCAGACGAACACTGGAAGATCATCTTAAAGAAAGAT-----TTAAGTCIC 2128
Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGlu----- 273
Db 2129 AATGATTTGGAG---CAACAAAAAATAAATTAAAGAAAGAT----- 289
Qy 274 ---AsnAlaGluLeuLysLysValLeuGlnMetLysLysGluMetIle----- 289
Db 2186 AATGAGGAAGACTTTGAAGCTGATAAGCAGATGGAAGAACCTTGCAATTCAGAAA 2245
Qy 289 ----- 289
Db 2246 CAGGTAGCAGAGAAACAGTTGAAGAAAGCAGAAAAATTGAATTGGAAGCAAGAAAAA 2305
Qy 289 ----- 289
Db 2306 ATAACCTGAAATTCAGTATACATGTAGAGAAAATGCAATTCGCAGTGTCCGATCACACAG 2365
Qy 290 -----SerLeuLeuSerProGlnLysLysProArgGluArgAlaGlu 304
Db 2366 GCTACATCATGCGAGCGAGTAACGGGTCTCCAGCAAGAACATGACAAAGCAGAAAGCAGAA 2425
Qy 305 Asp-----GlyThrGlyThrValAlaIleSerAspIleGluAspAspSer 319
Db 2426 GAACCTCAACAGCAGGTAGTGAACCTAAGCTGCCAATGAAGAGGCTGAACAGACATG 2485
Qy 320 GlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeu 339
Db 2486 AGAGAGCTGCATATGAA-----CTTAATGCCCTCCAGCTTGAAAAA 2527
Qy 340 ThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsn 359
Db 2528 ACGTCACTCAGGAAAAAGGCT---CGTTTGCTAAAGAGATAAACTAGATGAAACAAATAAT 2584
Qy 360 -----GlnAlaSerLysValHisSer 366
Db 2585 ACACCTCAGATGCCCTTAAGTTGGAGCTGGAAGGAAGGATCAGCGCGAGAAAGGGTATTCT 2644
Qy 367 GluGlyLeuAsnGlu-----GluAspVal 374
Db 2645 CAACAACTCAGAGAGCTTGTAGGCAATTTGAATCAAAACACACAGGTAAAGCTGAAGAAGCC 2704
Qy 375 IleSer-----ArgGlnAspHisGluGlnGluThrGluLys 386
Db 2705 ATGCAAGAAGCTAGTGATCTCAAGAAAAATAAAGCGCAATTATCAGTTAGAAATTAGAATCT 2764
Qy 387 LeuGluLeuGlu-----IleGluArgCysLysGlu 396
Db 2765 CTTAATCATGMAAAGGGAACCTACAAAGAGAAGTAGACAGAAATCAAGGGCAGCATGTCT 2824
Qy 397 MetIleLysAlaGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAsp 416
Db 2825 GTAGCTGAGAAGAATATTACGCAATTTAAATTCACAAATTCATTCTTTTCGAGATGAGAAA 2884
Qy 417 ThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLysGluArgLysGluGlu 436
Db 2885 GAATTAGAAAGACTACAAATCTGCCAG-----AGAAAAATCAGATCATCTAAAAAGAACAA 2938
Qy 437 TrpThrLeuPheLysGluGln---LysLysAsnPheGluArgGluArgSerPheThr 455
Db 2939 TTTGAGAAAAAGCCATGAGCAGTTGCTTTCAAAATATCAAAAGCTGAAAAAGAAAAAT---AAT 2995

Qy 456 GluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrp 475
Db 2996 GATAAAATCCAAAGGCTCAATGAAGAATTTGGAGAAA---AGTAATGAGTGTGCAGAGATG 3052
Qy 476 ValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLys 495
Db 3053 CTAATAACAAAAAGTAGAGAGCTTACT-----AGGCAGAAATTAATGAACCAAAATTA 3103
Qy 496 LeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgPro 515
Db 3104 ATGATGCAGAGAATTCAGGCAGAAATCA-----GAGAAATATAGTTTTAGAGAAAAAACT 3157
Qy 516 ArgGlnLysLysLeuHisSerVal-----AlaAsnGlyValProAlaCysThrSer 532
Db 3158 ATCCAGCAAAAGATGTGAAGCACTGMAAAATTCAGGCAGATGGTTTTAAAGATCAGTACGC 3217
Qy 533 LysLeuThrLysSerLeuProAlaSerProSerThr---SerAspPheArgGlnThrHis 551
Db 3218 AGCACAATGAACACTTGCATAACACAGACAAAAACAGACAGCAGATTTTCAAGAAAAAAT 3277
Qy 552 SerCysValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGluSerLys 571
Db 3278 AAATGCTAGAGAA-----GACCTGGCGAAAAAGTCAAAATTTG 3316
Qy 572 ProSerGluValAlaArgGluSerThrAspGlnLysTrpSerValGlnSer 588
Db 3317 GTAAGTGAATTTAAGCAAAAGTGTGACCAACACAGACATTATCATCCAGAAT 3367

Search completed: June 14, 2006, 01:52:55
Job time : 434 secs

THIS PAGE LEFT BLANK

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1258.8	46.8	2272	7	US-10-104-047-1497		Sequence 1497, Ap
2	1258.8	46.8	2272	16	US-11-072-512-1497		Sequence 1497, Ap
c 3	399.8	14.9	161874	15	US-11-121-086-75		Sequence 75, App
4	305.8	11.4	403	10	US-10-779-543-12377		Sequence 12377, A
c 5	130.6	4.9	513	7	US-10-029-386-5594		Sequence 5594, Ap
c 6	123.6	4.6	178	7	US-10-029-386-19350		Sequence 19350, A
c 7	61.2	2.3	30191	11	US-10-330-773-631		Sequence 631, App
c 8	60.4	2.2	621	10	US-10-450-763-26492		Sequence 26492, A
c 9	57.4	2.1	3750	8	US-10-437-963-34854		Sequence 34854, A
c 10	55.8	2.1	2093	7	US-10-104-047-1666		Sequence 1666, Ap
11	55.8	2.1	2093	16	US-11-072-512-1666		Sequence 1666, Ap
12	54.6	2.0	13407	10	US-10-737-318-27		Sequence 27, App
13	54	2.0	1199	8	US-10-425-114-10108		Sequence 10108, A
14	54	2.0	1567	8	US-10-424-593-27277		Sequence 27277, A
c 15	53.2	2.0	154817	7	US-10-085-117-334		Sequence 334, App
c 16	52.4	1.9	1062	9	US-10-767-795-5235		Sequence 5235, Ap
17	50.8	1.9	556	3	US-09-732-627A-536		Sequence 536, App

Qy	241	TGCTTCTGCA	CAGGAGAGAA	CATTGTGA	CAAAAGTATTTTCTTATCTTGTATCAGGAGCTGAC	300
Db	373	TGCCTTCTGCA	CAGAGATATA	TTGAACAGAGTATCTCATATCTTGTATCAGGAATTTGAC	432	
Qy	301	CACCTTCGGGTTCCTTCTGTATGAGANTCCAAAGTAAAGAGCCAAAGAGAGAATT	360			
Db	433	TACTTTTGGTTTTCCTTCATTATATGAAGAATCCAAAGGTAAAGAGACAAAGAGAGAGATT	492			
Qy	361	AAATATAGTCGCTGTTCTGAACCTGTATGAACAGCTGCTGTGCTTCAGCGGAAGAACCT	420			
Db	493	AAATATAGTAGCTGTATCTAAATTTGTAATGAGCTGCTTGTGCTTCAGCGGAAGAACCT	552			
Qy	421	GCTGGCCAGGAGCGGTGGAGACACAGAACTTGAAGCTGGGACAGTGAACATGGACCACT	480			
Db	553	TCTAGCTCAGGAAAATGTGGAGACACAGAAATTTGAAGCGGGAAGTATATGGACCATCT	612			
Qy	481	GCAGAGCTGTACGCCAACTTTAAGGAGCAGTTTGGAAACGCTCCAGGCGGGAGAGATGATCGG	540			
Db	613	ACAGAGCTGCTACTCAAACTTAAAGAACCAACTGGAAACCTCCAGGAGGGAAATGATTTGG	672			
Qy	541	GCTTCAAGAGAGACACAGGACGCTGCAGTGCAGAAACAGGAGTTTGTCACTCAGCTCCCTGAA	600			
Db	673	GCTTCAGGAAGACACACAGCAGTTACATGTGAAGACAGGNAATTTGCATCAGCTACTATAA	732			
Qy	601	GAATGAGAAAGATGAGGTACAAAATTA	660			
Db	733	GAATGAGAAAGATGAGGTGCAAAATATACAAAATATCATTGCAAGTCCGAGCTACTCAGTA	792			
Qy	661	TAATCATGATGTGAAGAGGAAGGCGTGAATATATAAGCTAAAGGAGCGCTGCATCA	720			
Db	793	TAATCATGATATGAAGAGAAAGAGCGGTGAATATATAAACTGAAGAAACGCTCTACATCA	852			
Qy	721	GCTCGTTATGAACAGAGAGGATAAAAAACATAGCCATGATGCTTTTAAATATGTGGGTGCG	780			
Db	853	ACTTGTATTATGAACCAAGAAAGATAAGAAATAGCTATGACNAATTTGAAATATGTCCGGAG	912			
Qy	781	AGCTGATGGCAACCGAGCTCATGGAGGACTGACAAAACAGAAACCCAGGAAATGAAGATGA	840			
Db	913	AGCTGATGGAAAAGAGGCTCCTCGAGGACTGGTAAACCTGAACCCAGGAATGAAGATGA	972			
Qy	841	GATGTACAAAATTCCTGTTGAATGATTTATGAGTACCGCCAGNAGCAGATCCTGATGGAGAA	900			
Db	973	AATGTATAAAAATTCCTTGTGAATGATTAATGAATATCGTCAGAAAACAAATCTCTAATTSAGAAA	1032			
Qy	901	CGCGGAGCTGAAGAGGTCTCTCCACGAGATGAAGAAGCAGAGATGATCTCTCTCTGTCTCC	960			
Db	1033	TGCAGAACTTAAGAAAGGTTCTTCCAAACAAATGAAAAGAAATGATTTCTCTCTCTCTCTCC	1092			
Qy	961	TCAGAAGAAGAACCCAGGGAAGAGCAGAGGACGCGCACAGGCACTGTGTCTATCTCCGA	1020			
Db	1093	CCAAAAGAGAAAACCTAGAGAAAAGTAGATGATAGTACAGAACTGT---	1149			
Qy	1021	TATAGAAGATGACTCTGGGGAACCTGAGCAGAGACAGCGGTGGGGCCCTTTCTGTGACAC	1080			
Db	1150	TGTTGAAGAAGATGCGGGGAACCTAAGCAGAGAGATATGTGGGAACCTTTCTGTGAAAC	1209			
Qy	1081	TGTCAGAGAGCAGCTGACAAACACATCAGAAAACAGTGGAGAAATTTTGAAGCTCATGT	1140			
Db	1210	TGTGAGAGAGCAGCTTACAAACAGACATCAGAAAACAGTGGAGAAATTTTGAAGTCAATGT	1269			
Qy	1141	AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGGCCCTTAATGAGGAGGACGT	1200			
Db	1270	AGAAAAGCTTGATAACCDAGTTTCAAGGTACACCTGGNAGGTTTAAATGATGAAGATGT	1329			
Qy	1201	CATCTCAGCAACAGCACTAGCAAGAGACTGAGAAAACCTGGAGCTGGAGATTGAGCGGTG	1260			
Db	1330	AATCTCAGCAACAGACCATGAACAAAGAAACCTGAAAACCTCGAGTTAGAAAATTCACGACGT	1389			
Qy	1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACACAGCAGCTGCGCCACCACTGTGA	1320			
Db	1390	TAAAGAAATGATTAATACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCTATATGA	1449			
Qy	1321	TGATGACACCACTCAGCTGTTGGCAGACTGTACTTGTGCTGAAGAAAAGGAACCCCTTAA	1380			

Db	1450	TGATGATACACCTTCACTATTACGAGACTGTTATTTGTTGGAAGAAAGGAAGCTCTCAA	1509
Qy	1381	AGAAGAGTGACCCCTTTTAAAGCAGCAAAAAGAAATTTTGAAGAGAGAAAGGCGAAGCTT	1440
Db	1510	AGAAGATGTCCTTTTAAAGAGCAGCAAAAAGAAATTTTGAAGGGAGAGACGAAGCTT	1569
Qy	1441	TACAGAAGCTGCCATTGCAATGGGGTTGGAG-----	1471
Db	1570	TACAGAAGCCGCTATTTCGCCTGGGATTGGAGATTGGCTTTCTCTGCCAAGTGTTTTACA	1629
Qy	1472	-----	1471
Db	1630	CAAGAAATCTTTTGGGTTATCCATCATCCACGGTCAGTAGTGTGGTGTCTGAGAAG	1689
Qy	1472	-----AGAAAGGGCTTTTGAAGAAGAGCGAG	1496
Db	1690	ACAGTCAACCAAGACTATATTTCGTGTTTAAAAAGAAAGGCAATTTGAAGAGAAAGAG	1749
Qy	1497	CCAGCTGGGTAAAGCAGCAGCTTTTAAACATGCGAACTTTGACACCGAGAACTCAGAAA	1556
Db	1750	CCAGTTGGTTAAAGCAGCAGCTTTCTAAATATGACTACCTTTTGACCACCGAGAACTCAGAAA	1809
Qy	1557	ATGTGAAACTTTTTCAGTGCCTTCTCAGGAAGTTCTGATCCAGACAATCTTTATAGTCCACT	1616
Db	1810	ATGTGAAACTTTTTCAGTGCCTTCTCAGGAAGTTCTGATTTGGGCAATCTTTATAGTGCAC	1869
Qy	1617	CACGGCCACGGCAAAAGAAAGCTACACAGTGTGCTTAATGGGGTGCACAGCTTTGCACATCAA	1676
Db	1870	CGAGGCGAGCGCAAAAGAGAGCTCACAGTGTGCTTAATGGGTCTCCAGTTTGCATGTCTA	1929
Qy	1677	AACTGACTAAATCTCTTTCCTGCCTCACCTTCTACTTTCAGACTTTTCGCCAGACACATTCAT	1736
Db	1930	AACTTACTAAATCTCTTTCCTGCCTTCACTTCCACTTCAGACTTTTGCACACAGCTTCT	1989
Qy	1737	GTGTGCTGGAACACAGTTTCCATCAGTGTGCTGTAATATAACTCTCTGAAGAAAGCTAAACCAA	1796
Db	1990	GCATATCTGAACATAGTTTCAATCAATGTACTGAATATAACTGTCTGAAGAAATTAACCAA	2049
Qy	1797	GTGAGGTTGCAAGAGAAAGCAGGATCAGAAAGTGGAGCGTGCAGTCAGGCGCCAGCTCGC	1856
Db	2050	ATCAGGTTGGAGAGAAAGCTACAATCAAAATGGAGTGTGGCGTCAGACCTGGATCAC	2109
Qy	1857	GGGAGGGTGCTTACAGCGGATGCTCCTCGCCCTTCAGGAGCGCTCACGGGGACCGAGATG	1916
Db	2110	AGGAAGTTGCTATAGTGGATGCTCCTTGAGCTACACAAATCTCATGTAGAAAAAGATG	2169
Qy	1917	ACTTACCTTAAATGTGGCGGCTCAGTGTGTTCCAGATGTGGCGCTAGAG-----	1968
Db	2170	ACTTACCTTAGACATGTGGACTGGAATTTTTTTTCATTAATGTGTTTCATCAAGTTTTCACAT	2229
Qy	1969	---AGTTGACACAGGGTGTAGCTAAAGTCAGTCGTCT	2003
Db	2230	CTAAGTTGAAACAGGGTGTGTCATAAAGTCAGTTATCT	2267

RESULT 2
US-11-072-512-1497
; Sequence 1497, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO

1677 AACTGACTAAATCTCTCCCTCAGCTTCTCAGACTTTTCGCGACACATTCAT 1736
1930 AACTTACTAAATCTCTCCCTCAGCTTCTCAGACTTTTCGCGACACATTCCT 1989
1737 GTGTGCTGAAACAGAGTTCCATCAGTGTCTGTAATACTCTGAAAGAAAGTAAACCAA 1796
1990 GCATATCTGAACATAGTTCAATCAATGACTGTAATAAATCTGCTGAAGAAATTAACCAA 2049
1797 GTGAGGTTGCAAGAGAAAGCAGGATCAGAGTGGAGGCTGAGTCCAGGCGCCAGCTCGC 1856
2050 ATCAGGTTGGAGGAGAACGTAACAAATCAAAATGGAGTGGCGCTCAAGACCTGGATCAC 2109
1857 GGGAGGGGTGCTACAGCGGATGCTCTCCGCTTCAGGAGCGCTCACGGGACCGAGATG 1916
2110 AGGAAGGTTGCTATAGTGGATGCTCTTGGAGCTACAAATTTCTCATGTGAGAAAAGATG 2169
1917 ACTTACCTTAAATGTGGGGGCTGAGTGTCTTCCAGAGTGTGGCTAGAGG----- 1968
2170 ACTTACCTTACATGTGGACTGGAATTTTTCATTAATGTGTTCATCAAGTTTCACAT 2229
1969 ---AGTTGACACAGGGGTGAGCATAAAAGTACGTCGTCT 2003
2230 CTAAGTTGAACAGGGGTGTGTATATAAAGTCAGTTATCT 2267

RESULT 3
US-11-121-086-75/c
; Sequence 75, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121.086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 75
; LENGTH: 161874
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-75

Query Match 14.9%; Score 399.8; DB 15; Length 161874;
Best Local Similarity 74.0%; Pred. No. 3.2e-98;
Matches 738; Conservative 0; Mismatches 202; Indels 57; Gaps 16;
1749 ACAGTTCCATCAGTGTGCTGTAATATACTCTGAAGAAAGTAAACCAAGTGAGGTTGCAA 1808
34291 ACAGTTCAATCAATGTACTGAATATACTGCTGAAGAAATTAACCAATCAGGTTGGAG 34232
1809 GAGAAAGCAGGATCAGAAAGTGGAGCGTGCAGTCCAGGCCAGCTCGCGGAGGGGTGCT 1868
34231 GAGAAATCAAAATGGAATGGAGTGGGATCAAGACCTGGATCAGAGAAAGTTGCT 34172
1869 ACAGGGATGCTCTCCGCTCTCAGGAGCGCTCAGGGGACCGGAGATGACTTACCTTAAA 1928
34171 ATAGTGGATGCTCTTGAGCTACACAAATTTCTCATGTAGAAAAGAGTACCTTAGA 34112
1929 TGTGGGGCTCAGTGTCTTCCAGATGTGCGGTAGAG-----AGTTGACAC 1977
34111 CATGTGGACTGGAATTTTTTTTCATTAATGTGTTCATCAAGTTTCACATCTAAGTTGAAC 34052
1978 AGGGGTAGCATAAAGTCAGTCGTC-----TAACTTAAGATGCTCAGAGTTGTTGTTTG 2032
34051 AGGGGTGCTATAAAGTCAGTTACTTAATACTTAAGATGGTCTGAGTTGTTGTTG 33992
2033 GACTTCGCTGTCTTCCCGGAGAGCTGAAATGCT-AAGCTACTTAAAGATGCAAGC 2091
33991 GACTTCCCTGTCTTCCCGGAGAGGTTGAAATCTTAAATCTTATTTAAAGGATATAAAG 33932

2092 TTTGG-----TTGTTGTGTAGTAAAGAGCCCTGGCTGTGCTGACTGCGAGAAATGCAT-G 2146
33931 CTTTGGATATGATTTTTTTAGTAAAGAGCATCTGGTTCTGTGAATAAAGGATGTATAG 33872
2147 GCCTTTGGATGAAACAGAGCGCTGGAATGA--TTGCTCCGCCAGGTACCGAGAAAGC 2204
33871 ATCTTTGGATGAAACAAAGACACTAGACTGAGTCTTCTTATAGGTAATTAAAGATAGC 33812
2205 ACTTTTAGGACTGGTTCCTGTAAACATTAATATTCGTCGCCAAGTGTGGTGGCATGG 2264
33811 ACTTTTAGGAAACTGATTAATGTTAAATGTTTAAATTTGTCCTCAATATAGTGGCATGG 33752
2265 AAGTGTAGCCCTTACTTGAATGATATCTGTAGATTTTAAACAAGCAGGTTCTTATATTTA 2324
33751 AAGTTAGCCCTTACTTGAATGATATCTGTAGATTTTAAACAAGCAGGTTCTTATATTTA 33692
2325 TTATGTTTGTGATTTTGGGATTAACCTCTCTTTCATATGTTT----- 2367
33691 TTATGTTTGTG-TTTGAAATTAACCTCTCTTTCATATGTTTAAATATAAGTGAAATTTA 33633
2368 ----GTGCTGTACATAAATATACATGACTATGTTAAGAGGCTTTAAGGTTTAAAAACTT 2423
33632 TGTATGTTTGTACATAGATACACATGATTTATGTTAAGAGGCTTTAAGATTTAAAAAGTTT 33573
2424 CACACATGCTTCAGTATAGCATTTTCATGCAATTTAAATGTTTTCAGTGGCATGGTGT 2483
33572 CACA-CAACCATAGTATAGTATTTTCATGCC-AGTAAATTTTTCAGTGGTATCTGTT 33515
2484 TACAGAGG--TTAGGACCACTGCCATGACAGTTAAGACTTTATTTTAAAGCCATCTGG 2541
33514 TACAGATGTTATTAGGACCAATTGATGCAATTAACATTTAAGA-ATTCTCTTTAATACATCTGG 33456
2542 GCAATAAAATTCAAAGCCCTTCAT-AAGCTGAGTTC---AGATACTAGAACTACTAA 2597
33455 GCAATAAATTTGAAAGGTTATTCATGAAGCTGAGTTCCTTAGAGTAATCAACACTACTAA 33396
2598 CGTTACATTTTTCAGATTTTAAAGCAATT-GTATTTTATTTTATATATATGT-GAATGTTAT 2655
33395 CATTACATTTTTCAGATTTTATGACATTAAGATTTTATTTTGTATATAGTATATAT 33336
2656 AATTTCTAAGAGGAATATTGATTATGGAGTAATGGGG 2692
33335 AATTTTAAAGGACTATTGATGATAGAAGATAGGG 33299

RESULT 4
US-10-779-543-12377
; Sequence 12377, Application US/10779543
; Publication No. US20050227917A1
; GENERAL INFORMATION:
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
; TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/076,555
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 09/217,471
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: 60/068,755
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/080,664
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/105,234
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 09/297,648
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: PCT/US99/01619
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: 60/072,910
; PRIOR FILING DATE: 1998-01-28

```
; PRIOR APPLICATION NUMBER: 60/075,954
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/080,114
; PRIOR FILING DATE: 1998-03-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12377
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-779-543-12377

Query Match      11.4%; Score 305.8; DB 10; Length 403;
Best Local Similarity 85.6%; Pred. No. 9.4e-74;
Matches 340; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 470 ATGACACCTCCAGAGCTGTACGCCAAACTTAAAGGAGCAGTTCGAAACGTCACAGCGG 529
Db 7 AGGGACCATCTACAGAGCTGTACTCAAACTTATGGAACTATGGAACTTCAAGAGG 66

QY 530 GAGATGATCGGGCTTCAAGAGAGACAGCAGGCTGCGAGTGCAGTGCAGAAACAGAGTTTGCAT 589
Db 67 GAAATGATTGGCTTCAGGAAAGACAGACAGTGTACAATGTAAAGACAGGAATTTGCAT 126

QY 590 CAGCTCTCGAAGATGAGAAAGATGAGGTACAAAATTACAAAATATCATAGCCAGCCGG 649
Db 127 CAGCTACTAAAGAATGAGAAAGATGAGGTGCAAAAATTCAAAAATATCATTCGAAGTCGA 186

QY 650 GCTACTCAGTATAATCATGATGTGAAGAGGAGGAGCGTGAATATAATAAGCTTAAGGAG 709
Db 187 GCTACTCAGTATAATCATGATGTGAAGAGGAGGAGCGTGAATATAATAAGCTTAAGGAG 246

QY 710 CGCTGTCATCAGCTCGTTATGAACAGAGGATGAATAAATACATAGCCATGGATGTTTTAAAT 769
Db 247 COTCTACATCAACTTGTATGAACAGAGAGTGAATAAATAGCTATGGACATTTTGAAT 306

QY 770 TATGTGGCTGAGCTGATGGCAACGAGGCTCATGAGAGCTGCACAAAACAGAACCCAGG 829
Db 307 TATGTGGGAGAGCTGTAGGAAAAAGAGGCTCTCTGGAGAGCTGGTAAAAACTGAAGCCAGG 366

QY 830 AATGAAGATGAGATGTACAAAATTCCTGTTGAATGATT 866
Db 367 AATGAGATGAATGTATAAAATTCCTGTTGAATGATT 403

RESULT 5
US-10-029-386-5594/c
; Sequence 5594, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 5594
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009265.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EST HUMAN HIT: BE694886.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q40554, EVALUE 1.20e+00
; OTHER INFORMATION: NT HIT: Q40554, EVALUE 1.20e+00

Query Match      4.6%; Score 123.6; DB 7; Length 178;
Best Local Similarity 80.9%; Pred. No. 3e-23;
Matches 144; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 829 GAATGAAGATGAGATGTACAAAATTCCTGTTGAATGATTATGATACCCAGAGACAGAT 888
Db 178 GAATGAAGATGAATGTATAAAATTCCTGTTGAATGATTATGATACCTCAGAACCAAT 119

QY 889 CCTGATGAGAACCCGAGCTGAAGAAGGTCCTCCAGCAGATGAAGAGGAGATGATCTC 948
Db 118 CCTAATGGAAAATGCAGAACTTAAAGAAGGTTCTTCAACAAATGAAAAAGGAAATGATTC 59

QY 949 TCTCTGCTCTCTCAGAAAGAGAGCCAGGAAAGAGCAGAGCAGCCAGGCACAGGCACT 1006
Db 58 TCTTCTTTCTTCTCCCAAAAGAGAAACCTAGAGAAAGAGTAGATGATGATACAGGAAT 1

RESULT 7
US-10-330-773-631/c
; Sequence 631, Application US/10330773
; Publication No. US20060040262A1
```

```
US-10-029-386-5594

Query Match      4.9%; Score 130.6; DB 7; Length 513;
Best Local Similarity 76.6%; Pred. No. 6.7e-25;
Matches 160; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 809 ACTGACAAAACAGAACCCAGGAAATGAAGATGAGATGTACAAAATTCCTGTTGAATGATTAT 868
Db 411 ACTGACAGTTCTTTCTCAGGAATGAAGATGAATGTATAAAATTCCTTGAATGATTAT 352

QY 869 GAGTACCCCGCAGAACGATCTGATGAGAACCGGAGCTGGAAGAAGTCTCTCCAGCAG 928
Db 351 GAATATCGTCAGAAACAAATCTTAATGGAAAATGCAGAACTTAAAGAGGTTCTTCAACAA 292

QY 929 ATGAAGAAGGAGATGATCTCTCTCTGTCCTCTCAGAAAGAAAGCCAGGAAAGAGCA 988
Db 291 ATGAAGAAGGAAATGATTTCTCTCTTCTCCCAAAAGAAACCTAGAGAAAGAGTA 232

QY 989 GAGAACGCCAGGCAGCTGTTGCTATCTC 1017
Db 231 GATGATAGTACAGGAACCTGTAAGTGGCTC 203

RESULT 6
US-10-029-386-19350/c
; Sequence 19350, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19350
; LENGTH: 178
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009265.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EST HUMAN HIT: AUI33150.1, EVALUE 1.00e-95
; OTHER INFORMATION: SWISSPROT HIT: P03654, EVALUE 1.10e+00
; OTHER INFORMATION: NT HIT: g114723825, EVALUE 8.00e-96
US-10-029-386-19350

Query Match      4.6%; Score 123.6; DB 7; Length 178;
Best Local Similarity 80.9%; Pred. No. 3e-23;
Matches 144; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 829 GAATGAAGATGAGATGTACAAAATTCCTGTTGAATGATTATGATACCCAGAGACAGAT 888
Db 178 GAATGAAGATGAATGTATAAAATTCCTGTTGAATGATTATGATACCTCAGAACCAAT 119

QY 889 CCTGATGAGAACCCGAGCTGAAGAAGGTCCTCCAGCAGATGAAGAGGAGATGATCTC 948
Db 118 CCTAATGGAAAATGCAGAACTTAAAGAAGGTTCTTCAACAAATGAAAAAGGAAATGATTC 59

QY 949 TCTCTGCTCTCTCAGAAAGAGAGCCAGGAAAGAGCAGAGCAGCCAGGCACAGGCACT 1006
Db 58 TCTTCTTTCTTCTCCCAAAAGAGAAACCTAGAGAAAGAGTAGATGATGATACAGGAAT 1

RESULT 7
US-10-330-773-631/c
; Sequence 631, Application US/10330773
; Publication No. US20060040262A1
```

```
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001300
; CURRENT APPLICATION NUMBER: US/10/330,773
; CURRENT FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 981
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 631
; LENGTH: 30191
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(30191)
; OTHER INFORMATION: n = A,T,C or G
US-10-330-773-631

Query Match      2.3%; Score 61.2; DB 11; Length 30191;
Best Local Similarity 43.7%; Pred. No. 0.0002;
Matches 413; Conservative 0; Mismatches 528; Indels 5; Gaps 3;

Qy 366 TAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTGCTGTTTCAGCGGAAGAACCTGCTGG 425
Db 14557 TTGAAGCTCGGCAGAGCAGGAGCAGGAGCAGGAGCAAGAGGAGGAGCAGGAGCAGGAGCAGG 14498

Qy 426 CCCAGGAGAGCTGGAGACACAGAACTTGAAGCTGGGCGAGTGACATGACCACTTCGAGA 485
Db 14497 AGCAGGAGGAAGACAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 14438

Qy 486 GCTGCTACGCCAAACTTAAGGAGCAGTTGGGAAACGCTCCAGGCGGAGATGATCGGC--- 542
Db 14437 AGGAGGAAGACGAGCAGCAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 14378

Qy 543 TTCAAGAGAGAGACAGGCGTGCTGAGTGCACAAAGAACAGGAGTTTGCATCAGCTCCTGAAGA 602
Db 14377 AGCAGGAGCAGGAGGAAGCAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGC 14318

Qy 603 ATGAGAAAGATGAGGTACAAAAATTAACAAATATCATGACCGCGGGCTACTCAGTATA 662
Db 14317 AGGAGGAGCAGGAGGAAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 14258

Qy 663 ATCATGATGTCAGAGGAGGAGGAGCGTGATATATAAGCTAAAGGAGCGCTGCATCAGC 722
Db 14257 AGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 14198

Qy 723 TCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTAAATATATGTGGTCCAG 782
Db 14197 AGGAGGAGCAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 14138

Qy 783 CTGATGGCAACAGGAGCTCATGGAGGACTGACAAAACAGAACCCAGGATGAAGATGAGA 842
Db 14137 AGGAGCAGGAGGAAGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 14078

Qy 843 TGTCAAAATTTCTGTTGAATGATTATGATACCGCCAGAACAGATCTGTATGAGAACG 902
Db 14077 AGCAGGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 14018

Qy 903 CGGAGCTGAAGAAGTCTCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTGTCCTC 962
Db 14017 AGGAGCAGGAGCTGGAGCAGGAGTAGGAGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 13958

Qy 963 AGAAGAGAACCCAGGGAAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1022
Db 13957 AGGAGCTGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 13898

Qy 1023 TAGAAGATGACTCTGGGGAACTG-AGCAGACAGACAGCGTGTGGGGCTTTCTGTGACACT 1081
Db 13897 AAGAGCAGGAGCTGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 13838

Qy 1082 CTGAGAGAGCAGCTGACAAACAGGATCAGGAAACAGTGGAGAAATTTTGAAGAATCATGTA 1141
Db 1141
```

RESULT 8

```
US-10-450-763-26492
; Sequence 26492, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 26492
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (269)..(595)
; OTHER INFORMATION: 35% homologous to Plasmodium falciparum glutamic acid-rich
; OTHER INFORMATION: protein, accession number AF251290, Smith-Waterman Score=193.
US-10-450-763-26492
```

```
Query Match      2.2%; Score 60.4; DB 10; Length 621;
Best Local Similarity 48.3%; Pred. No. 2.7e-05;
Matches 228; Conservative 0; Mismatches 241; Indels 3; Gaps 2;
```

```
Qy 503 AAGGAGCAGTTGGAAACGTCACGCGGAGATGATCGGCTTCAAGAGAGAGCAGGCAG 562
Db 103 AAGAGGAAGAGGAGGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 162

Qy 563 CTCAGTGCAGAACACAGGAGTTTGCATCAGCTCCTGAGATGAGAAAGATGAGGTACAA 622
Db 163 AAGAAGAGAGAGGAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 222

Qy 623 AAATTACAAATATCATAGCCGCTACTCAGTATATATCATGATGTGAAGAGGAAG 682
Db 223 AAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 282

Qy 683 GAGCGTGAATATATTAAGCTAAAGGAGCGCTGCATCAGCTCGTTATGAACAAGAGAT 742
Db 283 AAGGAGAG--AGAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 340

Qy 743 AAAAACTAGCCCTGGATGTTTAAATATATGTTGGTTCAGCTGATGCGAACAGGCTCA 802
Db 341 AAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 400

Qy 803 TGGAGGAGCTCAAAAAACAGAGCCAGGAATGAAGATCAGATGTACAAAAATTTCTGTGAAT 862
Db 401 GGGAGAAATCTAGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 460

Qy 863 GATTATGAGTACCCCCAGAGAGCAG-ATCCTGTGAGAGAACCGGAGCTGAAGAAGTCTCT 921
```


:	APPLICANT:	HIO, YURI
:	APPLICANT:	OTSUKA, KAORU
:	APPLICANT:	NAGAI, KEIICHI
:	APPLICANT:	IRIE, RYOTARO
:	APPLICANT:	TAMECHIKA, ICHIRO
:	APPLICANT:	SEKI, NAOHICO
:	APPLICANT:	YOSHIKAWA, TSUTOMU
:	APPLICANT:	OTSUKA, MOTOKYUKI
:	APPLICANT:	NAGAHARI, KENJI
:	APPLICANT:	MASUHO, YASUHIKO
:	TITLE OF INVENTION:	Novel full length cDNA
:	FILE REFERENCE:	084335-0191
:	CURRENT APPLICATION NUMBER:	US/11/072,512
:	CURRENT FILING DATE:	2005-03-07
:	PRIOR APPLICATION NUMBER:	US 60/350,978
:	PRIOR FILING DATE:	2002-01-25
:	PRIOR APPLICATION NUMBER:	JP 2001-379298
:	PRIOR FILING DATE:	2001-11-05
:	NUMBER OF SEQ ID NOS:	4096
:	SOFTWARE:	PatentIn Ver. 2.1
:	SEQ ID NO	1666
:	LENGTH:	2093
:	TYPE:	DNA
:	ORGANISM:	Homo sapiens
:	US-11-072-512-1666	
Query Match 2.1%; Score 55.8; DB 16; Length 2093;		
Best Local Similarity 44.6%; Pred. No. 0.0011;		
Matches 262; Conservative 0; Mismatches 322; Indels 3; Gaps 1;		
Qy	388	GAACGAGCTGCTCGTGTTCAGCGGGAAGAACCCTCGCCGCAGGAGAGCGTGGAACACA 447
Db	998	GGAGGAGAAGATGTGGAGACAGGAAGAAGCGTCCGAGCAGGAGAAAGNGCTGCCGGA 1057
Qy	448	GAACTTTGAAGCTGGGCGAGTGACATGGACCACTCTGCAGAGCTGCTACGCCAAAATTAAAGGA 507
Db	1058	GCAGGAAAGGAGCTGCGGGAGCAGAAAGAGCTSCGGGAGCAGGACAGATGCGAGGA 1117
Qy	508	GCAGTTGGAAACGTCACGGCGGGAGATGATCGGCTTCAGAGAGACACAGGCGACTGCA 567
Db	1118	GCAGGAGGAGAAGATGTGGGAGCAGGAGGAGAAGATCGGGAGCAGGAAGAAGATGTG 1177
Qy	568	GTGCAAGAACAGGAGATTGTGCATCAGCTCCTGAAGAATGAGAAAGATGAGGTACAAAAATT 627
Db	1178	GAGACAGGAGGAGAGGCTGTGGGAGCAGGAGAAGCAGATCGGAGCAGGAGCAGAAAGAT 1237
Qy	628	ACAAAATATCATACCGACCGGGCTACTCAGATATAATCATGATGTGAAGAGGAAAGGAGCG 687
Db	1238	CGCGGACCAAGAGGAGAGGATGTGGGAGCAGGACGAGAGGCTACCGGAGAGGAGGAGAG 1297
Qy	688	TGAATATAATAAGCTAAAGGAGCGCCCTGCATCAGCTCGTTATGAACAAGAGAGGATAAAA 747
Db	1298	GATCGGGAGCACAGAGAAGATGTG---GGACAGGTGGAGAAGATGCGGAGGAGAGAA 1354
Qy	748	CATAGCCATGGATGTTTTTAATATATGTGGTTCGAGCTGATGGCAAACGAGGCTCATGGAG 807
Db	1355	GATCGAGGAGCAGGAGAAGAAGACGCGGGACCCAGGAGGAGAAGATGCAAGAGGAGGAGAG 1414
Qy	808	GACTGACAAAACAGAACCCAGGATGAAGATGAGATGTACAAAATTCTGTTGAATGATTA 867
Db	1415	GATACGGGAGCGGAGAGAAGATGTCCGGGAAGAGGAGGAGACGATGCGGGAGCAGGAGGA 1474
Qy	868	TGAGTACCGCCAGAACGAGATCCTTGATGGAGAACCGCGAGCTGAAGAAGGTCCTCCACGA 927
Db	1475	GAAGATGCAGNAGCAGGAGGAGAAATATGTGGAGCACGAGAGAGAGNGNTGCGACGACGA 1534
Qy	928	GATGAAGAAGGAGATGATCT 974
Db	1535	GAGCTGCGGAAACAGAGGAGAGAAGCTGTGGGAACAGGACAGGAGAAGATGC 1591

; Sequence 27, Application US/10737318
 ; Publication No. US20050202442A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Morris, David W.
 ; APPLICANT: Malandro, Marc S.
 ; TITLE OF INVENTION: NOVEL THERAPEUTIC TARGETS IN CANCER
 ; FILE REFERENCE: 529452002800
 ; CURRENT APPLICATION NUMBER: US/10/737,318
 ; CURRENT FILING DATE: 2003-12-15
 ; NUMBER OF SEQ ID NOS: 92
 ; SOFTWARE: Fast-Seq for Windows Version 4.0
 ; SEQ ID NO 27
 ; LENGTH: 13407
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-10-737-318-27

 Query Match 2.0%; Score 54.6; DB 10; Length 13407;
 Best Local Similarity 48.8%; Pred. No. 0.0082;
 Matches 147; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

 Qy 544 AGCGGCTACTCAGTATATCATGTGTGAAGGAGGAGCGTGAATATATAAGCTA 703
 Db 7890 AGGAGGAGAAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7949

 Qy 704 AAGGAGCGCTGCATCAGCTCGTTATGAACAAGAAAGGATAAAAACATAGCCATGGATGTT 763
 Db 7950 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8009

 Qy 764 TTAATTATGTGGTTCGAGCTGTGCGCAACGAGGCTCATGGAGGACTGACAAACAGAA 823
 Db 8010 AAGAAGAAAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8069

 Qy 824 GCCAGGAATCAAGATGAGATGTACAAATTTCTGTTGAATGATTATGATGATACCGCAGAAG 883
 Db 8070 GAGNAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8129

 Qy 884 CAGATCTGATGGAGAAACGGCGAGCTGAAGAGGTCTCCAGCAGATGAAGAAGGAGATG 943
 Db 8130 AAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 8189

 Qy 944 A 944
 Db 8190 A 8190

 RESULT 13
 US-10-425-114-10108
 ; Sequence 10108, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E.
 ; APPLICANT: Tabaska, Jack E.
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
 ; FILE REFERENCE: 38-21(53313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 10108
 ; LENGTH: 1199
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: 700894116_FLI
 US-10-425-114-10108

Matches 190; Conservative 0; Mismatches 195; Indels 9; Gaps 1;
QY 589 TCAGTCTCTCAAGATGAGAAAGATCAGGTACACAAAATTTACAAAATATCATAGCCGCG 648
Db 142 TGAGAAAGCTGCAACAGGAGGGATGAATTCAGAGAAATGGTTATTTGGTAATCAGCAAGT 201
QY 649 GCGTACTCAGTATATCATGATGTGAAGAGGAGGAGCGTGAATATAATAAGCTAAAGGA 708
Db 202 AAAGACTCAACAATGCATGATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 261
QY 709 GCGCTGCATCAGCTGGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTTAAA 768
Db 262 GAGGCTAAACCAAGTGTTCATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 321
QY 769 TTATGGGTCGAGCTGATGGCAACAGGCGTCATGGAGGACTGACAAAACAGAGCCG 828
Db 322 GAATTTGCTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 380
QY 829 GAATGAAGATGAGATGTACAAAATTTCTGTTGAATGATTATGATGATACCGCCAGACGAT 888
Db 381 -----ATGATTTTATAAAGATTTGATGATGCTTATGATGATCAAAAAATCAAGAACT 432
QY 889 CCGTATGAGAGACGGGAGCTGAAGAAGGTCTCCAGCAGATGAAGAAGGAGGATGATCTC 948
Db 433 GATGCAGAGATGCTGATTTAAGAGCATATTAGATCAATGCAGGTGGATATCGGTGA 492
QY 949 TCTCTGTCTCTCAGAGAGAGAGCCGAGGAA 982
Db 493 TTTCTTAAATGCTCTTAATGGACTACCAAGCAA 526

RESULT 14
US-10-424-599-27277
; Sequence 27277, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 27277
; LENGTH: 1567
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_124632C.1
US-10-424-599-27277

Query Match 2.0%; Score 54; DB 8; Length 1567;
Best Local Similarity 48.2%; Pred. No. 0.003;
Matches 190; Conservative 0; Mismatches 195; Indels 9; Gaps 1;
QY 589 TCAGTCTCTCAAGATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCGCG 648
Db 508 TGAGAAAGCTGCAACAGGAGGGATGAATTCAGAGAAATGGTTATTTGGTAATCAGCAAGT 567
QY 649 GCGTACTCAGTATATCATGATGTGAAGAGGAGGCGTGAATATAATAAGCTAAAGGA 708
Db 568 AAAGACTCAACAATGCATGATGAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 627
QY 709 GCGCTGCATCAGCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTTAAA 768
Db 628 GAGGCTAAACCAAGTGTTCATGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 687
QY 769 TTATGGGTCGAGCTGATGGCAACAGGCGTCATGGAGGACTGACAAAACAGAGCCAG 828
Db 688 GAAATTTGCTTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 746

QY 829 GAATGAAGATGAGATGTACAAAATTTCTGTTGAATGATTATGATGATACCGCCAGAGCAGAT 888
Db 747 -----ATGATTTTATAAAGATTTGTTGATGCTTTATGATGATCAAAAAATCAAGAACT 798
QY 889 CCGTATGAGAGACGGGAGCTGAAGAAGGTCTCCAGCAGATGAAGAAGGAGGATGATCTC 948
Db 799 GATGCAGAGAAATGCTGATTTTAAGAGCAATTATTGAGATCAATGCGAGGTGGATATCGTGA 858
QY 949 TCTCTGTCTCTCAGAGAGAGAGCCGAGGAA 982
Db 859 TTTCTTAAATGCTCTTAATGGACTACCAAGCAA 892

RESULT 15
US-10-085-117-334/c
; Sequence 334, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 334
; LENGTH: 154817
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)---(154817)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-334

Query Match 2.0%; Score 53.2; DB 7; Length 154817;
Best Local Similarity 48.4%; Pred. No. 0.099;
Matches 148; Conservative 0; Mismatches 158; Indels 0; Gaps 0;
QY 640 AGCCAGCGGCGCTACTCAGTATATCATGATGTGAAGAGGAGGAGCGTGAATATAATAA 699
Db 49930 AGGCTCTCTCTTAATTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 49871
QY 700 GCTAAAGGAGCGCTGCATCAGCTCGTTATGAACAAGAGGATAAAAACATAGCCATGA 759
Db 49870 GAAGAAGAA 49811
QY 760 TGTTTTAAATTTATGTTGGTCCGAGCTGATGGCAACGAGGCTCATGGAGGACTGACAAAAC 819
Db 49810 GAAGAAG 49751
QY 820 AGAAGCCAGGAATGAAGATGAGATGTACAAAATTTCTGTTGAATGATTATGATGATCCGCA 879
Db 49750 AGGAG 49691
QY 880 GAACAGATCCTGATGAGAGAGCGGAGCTGAAGAGGTCTCCAGCAGATGAAGAAGGA 939
Db 49690 GGAG 949
QY 940 GATGAT 945
Db 49630 GAAGAT 49625

Search completed: June 14, 2006, 02:09:22
Job time : 2851 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2006, 23:04:18 ; Search time 71 Seconds
(without alignments)
4812.369 Million cell updates/sec

Title: US-10-644-084-1
Perfect score: 2692
Sequence: 1 cgtaggagagtgacaggagc.....ttgattatggagtaatgggg 2692

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 253354 seqs, 63461778 residues

Total number of hits satisfying chosen parameters: 506708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA New:*

- 1: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 2: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 3: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 4: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 5: /EMC_Celerra_SIDS3/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 6: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 7: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 8: /EMC_Celerra_SIDS3/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52.2	1.9	1497	6	US-10-953-349-6159 Sequence 6159, Ap
2	45.4	1.7	2479	7	US-11-293-697-1966 Sequence 1966, Ap
3	42.6	1.6	759	7	US-11-217-529-1065 Sequence 1065, Ap
4	40.2	1.5	1799	7	US-11-293-697-1885 Sequence 1885, Ap
5	38.8	1.4	1147	6	US-10-953-349-6314 Sequence 6314, Ap
6	38.8	1.4	3207	7	US-11-217-529-80932 Sequence 80932, A
7	38.8	1.4	5849	7	US-11-270-287-1 Sequence 1, Appl
8	38.4	1.4	11978	7	US-11-257-851A-63 Sequence 63, Appl
9	38.2	1.4	1046	6	US-10-953-349-23270 Sequence 23270, A
10	38	1.4	438	7	US-11-301-554-1028 Sequence 1028, A
11	37.4	1.4	1434	7	US-11-217-529-80879 Sequence 80879, A
12	37.2	1.4	898	6	US-10-953-349-36772 Sequence 36772, A
13	36.6	1.4	3495	7	US-11-217-529-5994 Sequence 5994, Ap
14	36.4	1.4	138941	6	US-10-489-730-10 GENERAL INFORMATI
15	36.4	1.4	1137	7	US-11-217-529-79782 Sequence 79782, A
16	36.4	1.4	1998	6	US-10-473-173A-3859 Sequence 3859, Ap
17	36.4	1.4	2782	6	US-10-473-173-29 Sequence 29, Appl
18	36.2	1.3	591	7	US-11-217-529-152 Sequence 152, App
19	36.2	1.3	1854	6	US-10-471-571A-441 Sequence 441, App
20	36	1.3	1158	7	US-11-217-529-1020 Sequence 1020, Ap
21	36	1.3	3161	7	US-11-293-697-1992 Sequence 1992, Ap
22	35.8	1.3	363	7	US-11-217-529-173450 Sequence 173450, A
23	35.8	1.3	795	7	US-11-217-529-77171 Sequence 77171, A
24	35.8	1.3	1287	6	US-10-471-571A-1659 Sequence 1659, Ap
25	35.4	1.3	534	7	US-11-217-529-82693 Sequence 82693, A

ALIGNMENTS

RESULT 1

US-10-953-349-6159 7 US-11-217-529-44 Sequence 44, Appl
; Sequence 6159, Application US/10953349 Sequence 121, App
; Publication No. US20060107345A1 Sequence 7843, Ap
; GENERAL INFORMATION: Sequence 5843, Ap
; APPLICANT: ALEXANDROV, Nickolai et al. Sequence 27, Appl
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES Sequence 25, Appl
; TITLE OF INVENTION: ENCODED THERBY Sequence 29021, A
; FILE REFERENCE: 2750-1579PUS2 Sequence 2132, Ap
; CURRENT APPLICATION NUMBER: US/10/953.349 Sequence 40050, A
; CURRENT FILING DATE: 2004-09-30 Sequence 76624, A
; NUMBER OF SEQ ID NOS: 40252 Sequence 7460, Ap
; SOFTWARE: PatentIn version 3.3 Sequence 7460, Ap
; SEQ ID NO 6159 Sequence 3, Appl
; LENGTH: 1497 Sequence 3, Appl
; TYPE: DNA Sequence 10075, A
; ORGANISM: Arabidopsis thaliana Sequence 32, Appl
US-10-953-349-6159 Sequence 77539, A
Sequence 1951, Ap
Sequence 17, Appl
Sequence 174036, A
Sequence 704, App

Query Match 1.9%; Score 52.2; DB 6; Length 1497;
Best Local Similarity 50.4%; Pred. No. 9.7e+05;
Matches 191; Conservative 0; Mismatches 173; Indels 15; Gaps 2;

Qy	586	GCATCAGCTCCTGAAGAATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAG	645
Db	522	GAATGAGAAGCTTCAAAAAGAGAGAGATGAGTTTCAACGAATGGTATTGCTAACCCAGCA	581
Qy	646	CCGGCTACTCAGTATATCATGATGTGAAGAGGAGCGGTGAATATATTAAGCTTAA	705
Db	582	AGTCAAGACCCACAATTATCATGAAACAAAAGAAAGAAAAGAGTACATAAAGTTACA	641
Qy	706	GGAGCGCTGCATCAGCTCGTTAT---GAACAAGAGGATAAAAACATAGCCATGGATGT	762
Db	642	GGAGAGGTAAACCAAGCTTTGATGGAGAAAAGAAAGAACGAGATCAGCCATGGAGAT	701
Qy	763	TTTAAATATTGTGGTTCGAGCTGATGCGCAACGAGGCTCATGGAGGAGCTGACAAAACAGA	822
Db	702	TATGAATCTACTCCAGAAAGAGGAGACACGCGTGGAAACATGGAGTGGGAAGAAAACCTGA	761
Qy	823	AGCCAGGAATGAAGATGAGATGTACAAAATTTCTTTGAATGATTATGATACCCCCAGAA	882
Db	762	CTC-----TGATTTCTACAAAAAAATTTGTGGATGCATATAGGCGGAAAAATCA	809
Qy	883	GCAGATCCTGATGGAGAACCGGAGGTGAAAGAGGTCTCCAGCAGATGAAGAGGAGAT	942
Db	810	AGAACTGATGGCAGAGACACTGATCTGAGAGCGTTACTAGCATCAACGCGGCTGACAT	869
Qy	943	GATCTCTCTCCTGCTCCT 961	

Qy 562 GCTGCGTCAAGAACAGAGTTTCATCAGCTCCTGAGATGAG 616
Db 1367 GGGGAGGAGGAGAGATGTGGGGCAGGAGAGATGTCGAGAGAGGAG 1421

RESULT 5
US-10-953-349-6314
; Sequence 6314, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6314
; LENGTH: 1147
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-6314

Query Match 1.4%; Score 38.8; DB 6; Length 1147;
Best Local Similarity 47.2%; Pred. No. 0.41;
Matches 152; Conservative 0; Mismatches 167; Indels 3; Gaps 1;

Qy 674 AAGAGGAAGGAGCGTGAATATAATAAGCTAAGAGGAGCGCTGCATCAGCTCGTTATGAAC 733
Db 549 AAGAGGAAGGAGGAGGAGAGAGAGGATAGAGAAAGATGGTAATACTCTGAAGAT 608

Qy 734 AAGAGGATAAAACATGACCATGATGTTTAAATATGTGGTFCGAGCTGATGGCAA 793
Db 609 GATGAGTTTAAAGAAAGAAAGAAAGAGCAGTACAAGGAGCATCATGATGATGAT 668

Qy 794 CGAGCTCATGAGGACTGACAAACAGACAGCCAGGATGAAGATGATGATGATGATGAT 853
Db 669 TATGATGAG 728

Qy 854 CTGTTGAATGATTATGAGTACCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 913
Db 729 AAGCATTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 785

Qy 914 AAGTCTCCAGCAGATGAGAGAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 973
Db 786 GATGATGAG 845

Qy 974 CCCAGGGAAGAGCAGAGGACG 995
Db 846 CACTATGATATGATGATGATG 867

RESULT 6
US-11-217-529-80932
; Sequence 80932, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 80932
; LENGTH: 3207
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-80932

Query Match 1.4%; Score 38.8; DB 7; Length 3207;
Best Local Similarity 56.2%; Pred. No. 0.76;
Matches 73; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 839 GAGATGTACAAAATTTCTTGAATGATTATGAGTACCGCCAGAACGACATCTCTGATGGAG 898
Db 839 GTGAAGGACAGATTTATTCAGAGGAGTAGAAATGGAATTAGAACGACATCTGATGGAG 898

Qy 899 AACGCGAGCTGAAGAAAGTCTCCAGCAGATGAAGAGGAGATGATCTCTCTCTCTCTCT 958
Db 899 AAAGCGATGCTGAAAAAGATGAAAGTCAGGATGAAGAGGAGAACATTCGGTAGATT 958

Qy 959 CCTCAGAGA 968
Db 959 TTCAAAATA 968

RESULT 7
US-11-270-287-1/c
; Sequence 1, Application US/11270287
; Publication No. US20060099627A1
; GENERAL INFORMATION:
; APPLICANT: Kara, Anna K.
; APPLICANT: Ting, Robert C.
; APPLICANT: Tham, Jill M.
; APPLICANT: Nelson, James S.
; APPLICANT: Tan, Theresa M.
; TITLE OF INVENTION: Diagnosis of Parasites
; FILE REFERENCE: 64-99
; CURRENT APPLICATION NUMBER: US/11/270,287
; CURRENT FILING DATE: 2005-11-09
; PRIOR APPLICATION NUMBER: US/09/369,992
; PRIOR FILING DATE: 1999-08-06
; PRIOR APPLICATION NUMBER: PCT/IB98/00212
; PRIOR FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: AU P09481/97
; PRIOR FILING DATE: 1997-09-26
; PRIOR APPLICATION NUMBER: AU P09329/97
; PRIOR FILING DATE: 1997-04-21
; PRIOR APPLICATION NUMBER: AU P04953/97
; PRIOR FILING DATE: 1997-02-06
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5849
; TYPE: DNA
; ORGANISM: Plasmodium berghei
US-11-270-287-1

Query Match 1.4%; Score 38.8; DB 7; Length 5849;
Best Local Similarity 58.8%; Pred. No. 1.1;
Matches 67; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 2572 TGAGTTCAGATACTAGAACTACTAAGCTTACATTTTGGAGATTTTAAAGCATTTGATT 2631
Db 726 TAATATAAAATTTATGATAATAATAATAATAATAATAATAATAATAATAATAATA 667

Qy 2632 TTATTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2685
Db 666 TTAATTAATAATATCAATATATATAATAATAATAATAATAATAATAATAATAATA 613

RESULT 8
US-11-257-851A-63/c
; Sequence 63, Application US/11257851A
; Publication No. US20060105951A1
; GENERAL INFORMATION:

```

; APPLICANT: Cunningham, Mark
; APPLICANT: Stojanovic-Susulic, Vedrana
; APPLICANT: O'Neill, Karyn
; APPLICANT: Huang, Chichi
; APPLICANT: Luo, Jeffrey
; TITLE OF INVENTION: Melanocortin Receptor Binding Mimetibodies, Compositions, Methods
; TITLE OF INVENTION: Uses
; FILE REFERENCE: CEN5080 USA NP
; CURRENT APPLICATION NUMBER: US/11/257,851A
; PRIOR FILING DATE: 2005-10-25
; PRIOR APPLICATION NUMBER: 60/621,960
; PRIOR FILING DATE: 2004-10-25
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 63
; LENGTH: 11978
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: In Vitro synthesized DNA with expression vector functions that
; OTHER INFORMATION: encodes an alpha-MSH mimetibody
US-11-257-851A-63

Query Match          1.4%; Score 38.4; DB 7; Length 11978;
Best Local Similarity 43.8%; Pred. No. 2.2; Mismatches 336; Indels 6; Gaps 2;
Matches 266; Conservative 0;

Qy 388 GAAAGAGCTGCTGCTTTCAGCGGAAGAACCTGTGCGCCAGGAGAGCGTGGAGACACA 447
Db 7070 GGAGCAGGAGCAGGAGCGGAGGGGCGGAGGACGAGGAGGGGCGAGGAGGAGGGGCA 7011

Qy 448 GAACTTGAAGCTGGGAGTGCATGACGACCACTCGAGAGCTGCTACGCCAACTTAAGGA 507
Db 7010 GGAGCAGGAGGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGGG 6951

Qy 508 GCAGTTGAAACGTCCTCAGGCGGGAGATGATCGGCTTCAAGAGAGACAGCGAGCTGCA 567
Db 6950 GCAGGAGCAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGCA 6894

Qy 568 GTGCAAGAACAGGAGTTTGCATCAGCTCTCTGAAGAATGAGAAAGATGAGGTACAAAAATT 627
Db 6893 GGAGCAGGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGG 6834

Qy 628 ACAAATATCATAGCCAGCCGGCTACTCATGATTAATCATGATGTAAGAGGAAGGAGCG 687
Db 6833 GCAGGAGCAGGAGGAGGGGCGAGGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAG-- 6776

Qy 688 TGAATATATAAGCTAAAGGAGCGCTGCATCAGCTCGTTATGAACAAGAGGATAAAAA 747
Db 6775 -GAGGGGCGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGCA 6717

Qy 748 CATAGCCATGGATGTTTTAAATTATGTGGTTCGAGCTGATCGCAAAACGAGGCTCATGGAG 807
Db 6716 GGAGGGGCGAGGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCA 6657

Qy 808 GACTGACAAAAACAGAGCCAGGAATGAAGATGAGATGTACAAAATTCTGTTGATGATTA 867
Db 6656 GGAGGGGCGAGGAGGAGGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGCA 6597

Qy 868 TGAGTACCGCCAGAGCAGATCCTGTATCGAGAAACGCGAGCTGAAGAGGTCTCTCCAGCA 927
Db 6596 GGAGGGGCGAGGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGA 6537

Qy 928 GATGAAGAAGGAGATGATCTCTCTCTGCTCTCTCCTCAGAGAAGAACGCCAGGGAAGAGC 987
Db 6536 GGGGCGAGGAGCAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGCA 6477

Qy 988 AGAGACG 995
Db 6476 GGAGCAGG 6469
```

```

RESULT 9
US-10-953-349-23270
; Sequence 23270, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 23270
; LENGTH: 1046
; TYPE: DNA
; ORGANISM: Glycine max
US-10-953-349-23270

Query Match          1.4%; Score 38.2; DB 6; Length 1046;
Best Local Similarity 54.7%; Pred. No. 0.57; Mismatches 0; Indels 0; Gaps 0;
Matches 76; Conservative 0;

Qy 825 CCAGGAATGAAGATGAGATGTACAAAATTCCTGTTGAATGATTATGAGTACCGCCAGAACG 884
Db 305 CCATGCTGTAATGGGAGAGAAACAAGAAGATGGCGAACCCAGGAGAGCTTCCTAGGGCATA 364

Qy 885 AGATCCTGATGGAGAACCGCGAGCTGAAGAAGTCTCTCCAGCAGATGAAGAAGGAGATGA 944
Db 365 TGATCGGAAGCAACAGGAGCAGCTGAAGAAGTCTGTGAAGGAAACAAGAGGAGGAGA 424

Qy 945 TCTCTCTCTCTCTCTCTCA 963
Db 425 TGACCGGTTCATGCATCA 443

RESULT 10
US-11-301-554-1028
; Sequence 1028, Application US/11301554
; Publication No. US20060088527A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Kalos, Michael D.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Durham, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C21
; CURRENT APPLICATION NUMBER: US/11/301,554
; CURRENT FILING DATE: 2005-12-13
; PRIOR APPLICATION NUMBER: US 10/283,017
; PRIOR FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: US 10/113,872
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 10/017,754
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 09/902,941
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 09/849,626
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US 09/736,457
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/702,705
; PRIOR FILING DATE: 2000-10-30
```

```

; PRIOR APPLICATION NUMBER: US 09/677,419
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 09/671,325
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 09/658,824
; PRIOR FILING DATE: 2000-09-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2157
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1028
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-301-554-1028

Query Match      1.4%;      Score 38;  DB 7;  Length 438;
Best Local Similarity 49.5%; Pred. No. 0.38;
Matches 98; Conservative 0; Mismatches 100; Indels 0; Gaps 0

Qy      591  AGCTCTCTGAAGAATGAGAAAGATGAGGTACAAAAATTACAAAATATCATAGCCAGCCGGG 650
Db      211  AGGACTTAAAGGAGAAGAAAGTGTGTGGAAGAGGCAGAAAATGGAAGAGACGCCCTG 270

Qy      651  CTACTCAGTATAATCATGATGTGAAGAGGAGGCGTGAATATATTAAGCTAAAGGAGC 710
Db      271  CTAACGGGAATGCTAATGAGAAAATGGGGAGCAGGAGGCTGCACAATGAGGTAGACGAAG 330

Qy      711  GCCTGTCATCAGCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGCTTTTAAATT 770
Db      331  AAGAGGAGAGAGGTGGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 390

Qy      771  ATGTGGCTCGAGCTGATG 788
Db      391  ATGGAGATGAAGATGAGG 408

```

```

RESULT 11
US-11-217-529-80879
; Sequence 80879, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 80879
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-80879

```

	Query Match	1.4%	Score 37.4	DB 7	Length 1434
	Best Local Similarity 52.2%	Pred. No. 1.1			
	Matches 83	Conservative 0	Mismatches 76	Indels 0	Gaps 0
Qy	2116	GCCTCGCTCTGTGACTGCAGGAATGCGTGGCGCTTTGGATGGAAACAGAAAGCGCTGGAA			2175
Db	652	GCCACACAGTGCCTTGAATGCAAGAAATGGTTCACCTTTTAAAGAAAAAGTATCGCAGAA			711
Qy	2176	TGATTGCTCGCCAGGTACCGAAGAGCAGCTTTTAGGACTGGTCTCTGTTAAACATTAA			2235
Db	712	TCCTTACATTCTCAACCAAGAAGACTTCGTCTTTTAGGGAATACATCAAGTACACAAGCA			771

```

Qy      2236 ATATTCTGCCCAAGTGTGGTTGTCATTTGGAAGTGTTAGCC 2274
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      772 AAAAAAGCGCAGRAGGCGGTTACGAGAGAAATCTTGCC 810

RESULT 12
US-10-953-349-36772
; Sequence 36772, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERAPY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 36772
; LENGTH: 898
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-36772

Query Match           1.4%; Score 37.2; DB 6; Length 898;
Best Local Similarity 61.2%; Pred. No. 0.97; Indels 0; Gaps 0;
Matches 60; Conservative 0; Mismatches 38;

Qy      877 CCAGAAGCAGATCCTGTATGAGAACGCGGAGCTGGAAGAAGTCTCTCCAGCAGATGAAGAA 936
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      357 CCGGCTCCCATGATGATGTCGTACGGCCGCAGAACGAGCACTTGCTGCAGGNACAGAC 416

Qy      937 GGAGATGATCTCTCTCTGTCTCTCTCAGAAAGAAGAC 974
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      417 ACTGAACAAGCGCTCGTCCGGCTCTCTGAGAAGAGGCAGC 454

```

```

RESULT 13
US-11-217-529-5994
; Sequence 5994, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5994
; LENGTH: 3495
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-5994

```

	Query Match	1.4%	Score 36.6;	DB 7;	Length 3495;
	Best Local Similarity	46.6%;	Pred. No. 3.3;		
	Matches 117;	Conservative 0;	Mismatches 134;	Indels 0;	Gaps 0;
Qy	543	TTCAAGAGAGACAGCGCAGCTGCACTGCAAGAACAGGAGTTTGCATCTCAGCTCCTGAAGA	602		
Db	3089	TTGAGAAAGAAAGAAAGAGGAGAGCGAAGATAAGAAAGAAAGAAAGAGGAGGAAGA	3148		
Qy	603	ATGAGAAAGATGGGTACAAAAAATTCAAAANTATCTAGCCAGCCGGCTACTCTAGTATA	652		
Db	3149	AGAAAGAAAGAGGAGGAAGAGGAGGAGAGAGAAAGAAAGAGGAGGAAGAGGAAAGAGAAAGAA	3208		

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 12, 2006, 19:14:14 ; Search time 42 Seconds
(without alignments)
1408.888 Million cell updates/sec

Title: US-10-644-084-2
Perfect score: 3165
Sequence: 1 MGDWMTVPVLCTENKNLS.....CYSGCSAFRSHGDRDDL 615
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2786.5	88.0	614	2	T42649
2	238.5	7.5	1738	2	T14867
3	237	7.5	2331	2	T25410
4	234.5	7.4	1679	2	S48385
5	224	7.1	2139	2	T18296
6	223.5	7.1	3459	1	A58539
7	222.5	7.0	3225	2	I52300
8	220.5	7.0	1780	2	T17272
9	220.5	7.0	1875	2	S38173
10	220	7.0	2442	2	T08621
11	218.5	6.9	1128	2	G82666
12	218.5	6.9	1354	2	S74244
13	218.5	6.9	1538	2	T25095
14	217.5	6.9	746	2	T47237
15	217.5	6.9	1909	2	A45592
16	215.5	6.8	1388	2	S74245
17	215.5	6.8	2649	2	A40937
18	214	6.8	980	2	F16066
19	214	6.8	1085	2	E96712
20	213	6.7	944	2	S26710
21	213	6.7	1955	2	T30934
22	212.5	6.7	2653	1	S28261
23	212	6.7	1790	2	S67593
24	212	6.7	1979	1	S03166
25	211.5	6.7	1181	2	T01799
26	211.5	6.7	1388	2	S70633
27	211	6.7	2253	2	T30336
28	210.5	6.7	1354	2	S69211
29	210	6.6	911	2	SS1441

ALIGNMENTS

RESULT 1

T42649
hypothetical protein DKFp434C0515.1 - human
C:Species: Homo sapiens (man)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T42649
R:Blöcker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, November 1999
A:Reference number: Z22230
A:Accession: T42649
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-614 <AAA>
A:Cross-references: UNIPROT:O9UIX0; UNIPARC:UP1000006DABF; EMBL:AL133046
A:Experimental source: adult testis; clone DKFp434C0515
C:Genetics:
A>Note: DKFp434C0515.1

Query Match		88.0%;	Score 2786.5;	DB 2;	Length 614;
Best Local Similarity		87.5%;	Pred. No. 3e-127;		
Matches 538;		Conservative 35;	Mismatches 41;	Indels 1;	Gaps 1;
QY	1	MGDMWMTVPVLCTENKNLSQVTSKMPSSLYSQVLCSSVPLSKNVHGVFGVCTGE	60		
DB	1	MGDMWMTVPGLSSSKTISQVTSKMPSSLYSQVLCSSVPLSKNVHSPFSAFCTD	60		
QY	61	NIEQGISVLDQELTTFGPSPSYEESKSKAKRELNIIVAVLNCWELLVLQKNLLAQESV	120		
DB	61	NIEQGISVLDQELATTFGPSPSYEESKSKETKRELNIIVAVLNCWELLVLQKNLLAQENV	120		
QY	121	ETQNLKLGSDMDHLQSCYAKLKEQLTSRRREMIGLQERDROLQCKNRSVLHQLLKNKDEV	180		
DB	121	ETQNLKLGSDMDHLQSCYAKLKEQLTSRRREMIGLQERDROLQCKNRLHQLLKNKDEV	180		
QY	181	QKLNIIASRATQYNHDKRKEREYNKLERHLQHLVNMKKDKNIAMDVLNYYVGRADGKRG	240		
DB	181	QKLNIIASRATQYNHDKRKEREYNKLERHLQHLVNMKKDKKIAMDILNYYVGRADGKRG	240		
QY	241	SWRTDKTEARNEDEMYKILLNDYEYRQKQILMENAEKLVLOOMKEMISLLSPQKKPR	300		
DB	241	SWRTDKTEARNEDEMYKILLNDYEYRQKQILMENAEKLVLOOMKEMISLLSPQKKPR	300		
QY	301	EAEQDGTGTVASDIEDSGELSRDSVGLSCDTVREQLTNSIRKQWRLKSHVEKLDNQ	360		
DB	301	ERVDDSTGTV-ISDVEEDAGELSRDSVGLSCDTVREQLTNSIRKQWRLKSHVEKLDNQ	359		
QY	361	ASKVHSEGNBEDVTSRDHEQETEKLEIERCKEMIKAQOQLLQOQLATTCDDDTTSL	420		
DB	360	VSKVHLEGFNEDVTSRDHEQETEKLEIEIQCKEMIKTQOQLLQOQLATAYDDTTSL	419		
QY	421	LRDCVLLBEKRLKEWTLFKQCKNFRERRRSFTAAAIRGLERKATFEERASWVKQOF	480		

hypothetical prote
hypothetical prote
364K Golgi complex
tpr protein - huma
nuclear matrix con
microtubule-vesicl
hypothetical coile
myosin ii - fissio
hypothetical prote
probable nuclear p
resin - human
hypothetical prote
serine/threonine p
early endosome ant
protein F4IH10.4 [I
microtubule bindin

Db 2292 SRISTA 2297

RESULT 4
S48385
hypothetical protein YIL149c - Yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 09-Jul-2004
C:Accession: S48385
R:Churcher, C.
submitted to the EMBL Data Library, September 1994
A:Reference number: S48310
A:Accession: S48385
A:Molecule type: DNA
A:Residues: 1-1679 <CHU>
A:Cross-references: UNIPROT:P40457; UNIPARC:UPI000013B453; GB:Z47047; EMBL:Z38059; NID:9
A:Gene: SGD:MIP2; MIPS:YIL149C
A:Cross-references: SGD:S0001411
A:Map position: 9L

Query Match 7.4%; Score 234.5; DB 2; Length 1679;
Best Local Similarity 19.6%; Pred. No. 0.00087;
Matches 145; Conservative 124; Mismatches 230; Indels 241; Gaps 28;

QY 15 ENKLSQYT-----SETKSPSSLYSQVLCSSVPLSKNVHGVFGVCTGE--- 60
Db 1010 ENKJSELVIRLEKDAADCAELTKTKSSLYSAQDL-----LDKHKRWMEEKADYEREL 1064

QY 61 --NIEQS-----ISYLDQELTFG-----PPSLYBESKSEAKRELNIVA 98
Db 1065 ISNIEQTESLRVENSVLIEKVDDTAANGDKHKLVLFSNLRHNSLETK----- 1117

QY 99 VLNCMNELLVLRKNLLAQESV--ETQNLKLGSDMDHLOSCYAKLKBQLETSRREMIG--- 154
Db 1118 LTTCKRELAFVKQKNDSTINDLQRTQTLSEKEY--QCSAVIIDEFKDITKEVTQVNI 1175

QY 155 -----LQERDROCKVRSLSHLLKNEKDEQVKLONIIASRATQY---NHDVKKERE 204
Db 1176 LKENNAILOKSLKNVTEKREIYKQLNDRQEETSRLQDLIQTKQVSNKILVYSE 1235

QY 205 YNKLKERLHQLVMNKKNIAMDVLNVYGRADKRGSRWTDKTEARNEDEMYKILLNDYE 264
Db 1236 MEQCKORYDLSSQQKD-----AQKDIETKLTNE-----ISDLK 1269

QY 265 YRQKQILMENAEKLVLOQMKKEMISLSLPQKKPRERAEDGTGTVAISDIEDDSGELSR 324
Db 1270 GKLSAENANADLENKFNLLKQQAHEKLDASKKQ----- 1303

QY 325 DSVWGLSCDTVREQLTNSRQKWRILKSHVEKLDNQA-----SKVHSEGLNEED 373
Db 1304 -----QAALTNELNE--LKAIKDKLEQDLHFENAKVIDLDLTKLKAHELQSD 1348

QY 374 VISRODHEOET-----EKLELETERCKE-MIKAAQOOL 404
Db 1349 V--SRDHEKQDRTLMEEITESLKKELQIFKTANSSSDAFEKLVNKEKDRIIDERTKE 1406

QY 405 LQOQLATTCDDDTSLRLDCYLLEEKERLKEEW-----TL--PKEQKKNFERRRSFT 455
Db 1407 FEKKLQETLNKSTSS--BAEYSKDIETLKKEWLKEYEDETLLRIKEAENLKKRILRPS 1463

QY 456 EAAIR--LGLERKAFEEERASVWQQ----- 479
Db 1464 EERIQKIIISKRKEELEEFKRLKENAGSLTFLDNKSGEDAEELWNSSPSKGNRPSA 1523

QY 480 ---FLNMTNFDHQS--ENVKLFSAFGS-----SDPDNLI VHSPRQKKLHVANGVPAC 530
Db 1524 VAGFINQKMLKQEQBLKVNNDVFNDSQMTNKENNIVDSAAKKAIPTFSGKPPF 1583

QY 531 TSKLTKSLPASPTSDFRTHSCVSHSSISVINITPTE-----SKPSEVARESTDQ 582
Db 1584 SSN-TSSLQS--FQNPFTASQSNINTNAPLRLTINTLOPEVAVKAAINFSNVTDLTNNSTDG 1640

QY 583 KMSVQSRPSSREGCYSGCSS 602
Db 1641 AKITEIGSTSKRPIESGTSS 1660

RESULT 5
T18296
myosin heavy chain - Entamoeba histolytica
C:Species: Entamoeba histolytica
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18296
R:Guillen, N.
submitted to the EMBL Data Library, February 1997
A:Reference number: Z18865
A:Accession: T18296
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2139 <GUI>
A:Cross-references: UNIPROT:Q07569; UNIPARC:UPI00000802D3; EMBL:L03534; NID:9L850912; P
C:Genetics:
A:Gene: mhca
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:91-780/Domain: myosin motor domain homology <MMO>

Query Match 7.1%; Score 224; DB 2; Length 2139;
Best Local Similarity 20.5%; Pred. No. 0.0037;
Matches 158; Conservative 110; Mismatches 240; Indels 262; Gaps 32;

QY 15 ENKLSQYTSETKMPSSLYSQVLCSSVPLSKNVHGVFGVF-----CTGE--NIE 63
Db 1100 EBDTLSQNLNLEKLTNTK-----TKADLEKKISGLQDYEDLEDDKKNKEGDLRNAQ 1154

QY 64 QSISYLDQELTTFGFPPLY-----BESKSEAKRELNIVAVLNCMNELLVLRKNLLAQE 118
Db 1155 RKIKELDDIITGADVQSVYLQKQEEYSQIAKMQEKEAIGND-----VKNKEKTIKEK 1209

QY 119 SVETQNLKLGSDMDHLOSCYA-KLKBQLETSRREMIGLQERDROCKVRSLSH-----QLL 173
Db 1210 ELEIQLSQLEKLDTEVEKEDAEKKKEIE--KEMKALQEEKENVENESSNTEKDKKKLE 1266

QY 174 KNEKDEQVKLONIIAS-----RATQVNHQV-----KPK--BREVN 206
Db 1267 DNLKQTKQLDDMTADNEKLLKAKADKLEAQLNEVQNHEKAVADAEKLNKKAQSKELN 1326

QY 207 KLKERLHQLV-----MKNKDKNIAMDVLNVYGR----- 234
Db 1327 SLKAELEALTAKSVESKKNKDSENEKAALSEEIDQANEKKNIQADLRKATADLQEAENE 1386

QY 235 -----ADKRGSRWTDKTEARNEDEMYKILLNDYE----- 264
Db 1387 KKAERVEAQRDKLVADNKKMTKTLEBIKARDEENTYKV--ENYEKVLKRKEADLEAEANL 1444

QY 265 -----YRQKQILMENAEKLV-----LQOMKK----- 286
Db 1445 DIEKQDRNNKEQVKLEGELETKDKLNAATAEKDSIFTAKKQSDADLEELNKTVEEHD 1504

QY 287 EMISLLSPQ-----KKKPRERAED-----GTGT--- 309
Db 1505 EVVAKLNTQITKLTRDNQSAEELNELRSKADKKKISELEEQVNELESRPVGTGNAD 1564

QY 310 --VAISD--IEDDSGELSRDSVWGLSCDTVRQL---TNSIRKQWRILKSHVEKLDN--- 359
Db 1565 NEIKIRDAQIADNLKALEMKGQVNNQLQATNKLKAKNDLTSKIEITENEMKKLENAKK 1624

QY 360 --QASKVHSEGLNEEDVLSRDHEOETKLELEIERCKEMIKAAQOOLLQOQLATTCDDDT 417
Db 1625 RLEQDKDADKAVSQTIKRGLEBEEVKKLTTEIQALKFQINAPSSVAQE----- 1674

QY 418 TSLLRDCYLLBEKERLKEEWTLFKQKKNFERRRSFTAAIRLGLERKAFEEERASVWK 477
Db 1675 -----BEKORLESIDIAELKEQ---LEQERTTAANAE-----AERKKIQAE-LDEVK 1716

QY 478 QQFLNMTN-----FDHQNSENKVLPSAFSGSSDPDNLIVHSPRQKKLHVANGVFACTSK 533

Db 1717 FNLEVDVNTQREKLVAKNSN-----DAEIDSL-----KEEKALEDETEKITDDNNK 1763

QY 534 LTKSLPASSTSDPFRQTHSCVSEHSSISVLNITPESKPSVARESTDOK 583

Db 1764 LSEEI-----DSLDRKYNALLDSKSDVSMKERFQDELKVTKDALE-TEKK 1808

RESULT 6

A56539

giantin - human

N:Alternate names: macrogolgin

C:Species: Homo sapiens (man)

C:Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 09-Jul-2004

C:Accession: A56539; S57536

R:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M. Mol. Cell. Biol. 14, 2564-2576, 1994

A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein (A56539; MUID:94187728; PMID:7511208

A:Accession: A56539

A:Molecule type: mRNA

A:Residues: 1-3259 <SSE>

A:Cross-references: UNIPROT:Q14789; UNIPARC:UPI000012B42C; EMBL:X75304; NID:9405714; PID F:3238-3254/Domain: transmembrane #status predicted <TMN>

C:Genetics:

A:Gene: GDB:COLGB1; GCP: GCP371

A:Cross-references: GDB:454958

A:Map position: 3q13.31-3q13.31

C:Superfamily: giantin

C:Keywords: coiled coil; Golgi apparatus; transmembrane protein

F:3238-3254/Domain: transmembrane #status predicted <TMN>

Query Match 7.1%; Score 223.5; DB 1; Length 3259;

Best Local Similarity 19.9%; Pred. No. 0.0065;

Matches 147; Conservative 136; Mismatches 218; Indels 237; Gaps 34;

QY 15 ENKNLSQYTSFKMS-----PSSLYSQVLCSSVP-LSKNVHGVFGVFCGTENIEQ 64

Db 1225 ENENIGDQLRQLQVRESIDGKLPT--DQESCSSTPGLEELFKATEQHHTQPVLES 1282

QY 65 SI-----SYLDELTTFGPFSLYE---ESKSEKAERELNIVAVLNCMELLVLQRKNLLA 116

Db 1283 NLCPWPMSHSDASALQGGTSVAQIKAQKLEIAEKVELEKVSSTTSELTKKSEVPQL 1342

QY 117 QESVETQNLKLS--DMHLOSCYAK-LKEQLETSRRRMIGLQERDROLQCKNRSLHQLL 173

Db 1343 QEQINKQGLEIESLKTVSHEAEVHAESLQOKLESSQLQIAGL-EHLRELQPK----- 1393

QY 174 KNEDEVQKQNIIASRATQYNHDKREREYNKLERHLQVLMNKKDKNIAMDVLNVYG 233

Db 1394 ---LDELQKL-----ISKKEEDVSYLGGQ-----SEKAALTKI----- 1425

QY 234 RADKRGSRWTDKTEARNEDEMYKIL-----LNDYEVROKQILMENAELK----- 278

Db 1426 -----QTEIIEQEDLIKALHTQLEMQAKEHDERIKQLOVELCEMKQKPEEIGE 1473

QY 279 --KVLQMKKEMISLLSPQK-----KKPRERAEDGTGV-----AISDIED--DSGELS 323

Db 1474 ESRKQIQIRKQQAALISRKEALKENKSLQBELSLARGTIERLTSLADVESQVSAQNK 1533

QY 324 RDSVWG-----LSCDTVR-----BOL 339

Db 1534 KDTVLGRLLAQERDKLITEMDRSLLENQSLSSCSLKLALGLEGTEDKEKLVKEIESL 1593

QY 340 TNS-----IRKQWRILKSHVEKLDNQASKVHS--EGLNEE-----DVLRS 377

Db 1594 KSSKIASTEWQEKHKELOKEYEIILLOSVENVSNAEARIQHVVEAVROEKQELYGKJRS 1653

QY 378 ODHEQETEK-----LELEIERCKEMIK-----AQOQLLOQ-----LATTC 413

Db 1654 EANKKETEKQLOAEQAEQMEENKMKRFAKSKQKILEBENDRLRAEVHPAGDTAKEC 1713

QY 414 DDDTTSLLRDVYLEEKERLKEWTLFKQKQKNFRERRSFTEAARIGLERKAFEBERA 473

Db 1714 ME--TLISSNASMKEELERVRMEYETLSKFKQSLMSEKDSLSEEVQDL-----KHQIEDNV 1767

QY 474 SNAVQOFLNMTNFDHQNSNVK---LFSAFSGSSDPDNLIVHSRPRQKKLHSHVANGVPAC 530

Db 1768 S--KOANLEATE-KHDNQTNTVEGTQIPETEEQDSLMSSTR-----PTC 1811

QY 531 TSKLTKSLPASSTSDPFRQTHSCVSEHSSIS-----VLNITPESKPSVARE 577

Db 1812 SESVPSAKSANPAVKOPF-----SSHDEINNVLIQDQLKERIAGLEEKQKNKFSQ 1864

QY 578 ESTDQKSVQSRPSREG 595

Db 1865 TLENEKNTLLSQISTKDG 1882

RESULT 7

152300

giantin - human

N:Alternate names: gcp372

C:Species: Homo sapiens (man)

C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 09-Jul-2004

C:Accession: 152300

R:Sonoda, M.; Misumi, Y.; Fujiwara, T.; Nishioaka, M.; Ikehara, Y. Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994

A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in A:Reference number: 152300; MUID:95100974; PMID:7802676

A:Accession: 152300

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-3225 <RES>

A:Cross-references: UNIPROT:Q14789; UNIPARC:UPI000016A987; GB:D25542; NID:9662389; PIDN C:Superfamily: giantin

Query Match 7.0%; Score 222.5; DB 2; Length 3225;

Best Local Similarity 18.9%; Pred. No. 0.0071;

Matches 139; Conservative 147; Mismatches 217; Indels 233; Gaps 32;

QY 15 ENKNLSQYTSFKMS-----PSSLYSQVLCSSVP-LSKNVHGVFGVFCGTENIEQ 64

Db 1191 ENENIGDQLRQLQVRESIDGKLPT--DQESCSSTPGLEELFKATEQHHTQPVLES 1248

QY 65 SI-----SYLDELTTFGPFSLYE---ESKSEKAERELNIVAVLNCMELLVLQRKNLLA 116

Db 1249 NLCPWPMSHSDASALQGGTSVAQIKAQKLEIAEKVELEKVSSTTSELTKKSEVPQL 1308

QY 117 QESVETQNLKLS--DMHLOSCYAK-LKEQLETSRRRMIGLQERDROLQCKNRSLHQLL 173

Db 1309 QEQINKQGLEIESLKTVSHEAEVHAESLQOKLESSQLQIAGL-EHLRELQPK----- 1359

QY 174 KNEDEVQKQNIIASRATQYNHDKREREYNKLERHLQVLMNKKDKNIAMDVLNVYG 233

Db 1360 ---LDELQKL-----ISKKEEDVSYLGGQ-----SEKAALTKI----- 1391

QY 234 RADKRGSRWTDKTEARNEDEMYKIL-----LNDYEVROKQILMENAELK----- 278

Db 1392 -----QTEIIEQEDLIKALHTQLEMQAKEHDERIKQLOVELCEMKQKPEEIGE 1439

QY 279 --KVLQMKKEMISLLSPQK-----KKPRERAEDGTGV-----AISDIED--DSGELS 323

Db 1440 ESRKQIQIRKQQAALISRKEALKENKSLQBELSLARGTIERLTSLADVESQVSAQNK 1499

QY 324 RDSVWG-----LSCDTVR-----BOL 339

Db 1500 KDTVLGRLLAQERDKLITEMDRSLLENQSLSSCSLKLALGLEGTEDKEKLVKEIESL 1559

QY 340 TNS-----IRKQWRILKSHVEKLDNQASKVHS--EGLNEE-----DVLRS 377

Db 1560 KSSKIASTEWQEKHKELOKEYEIILLOSVENVSNAEARIQHVVEAVROEKQELYGKJRS 1619

QY 378 ODHEQETEK-----LELEIERCKEMIK-----AQOQLLOQ-----LATTC 413

Db 1620 EANKKETEKQLOAEQAEQMEENKMKRFAKSKQKILEBENDRLRAEVHPAGDTAKEC 1679

A;Map position: 11R

Query Match 7.0%; Score 220.5; DB 2; Length 1875;

Best Local Similarity 20.2%; Pred. No. 0.0047; Mismatches 119; Gaps 27;

Matches 134; Conservative 119; Indels 171; Gaps 27;

QY 16 NKNLSQYTSSTKSPSSLYSQ-QVLCSS---VPLSKNVHGVGVFCT-----58

DB 1107 NSRIEDLSQNKL-----LYDQIQIYTAADKEVNVNSTNGPLANNILTLRRERDILDTKVT 1162

QY 59 -----GENTEQISYLDQELTTFGFSPSLYEESKSK-----BAKRELNIIVA-----VLNOM 103

DB 1163 VAERDAKMLRQKISLMDVEL-----QDARTKLDNSRVEKENHSHIIQQHDDIMEKL 1213

QY 104 NELLVLQRNLLAQESVETONLK---LGSMDHLQSCYAKLKEQLETSRPMIGLQERDR 160

DB 1214 NQNLRLRESNITLRNELENNNNKKELQSELQKQNVAPISELTALKYSM---QEKBEQ 1270

QY 161 QLOCKNRSIHLQKNEKDEVOQLQNIIASRATQYNHDKRREYNKLERHLQVLVMMKK 220

DB 1271 ELKLAKEEVHRWKKRSQDILEKHEQLSSS-----DYEKLESEIENLKEELE---NKE 1319

QY 221 DKNI-AMOVNLVYGRADGKRGSWRTDKTEARNEDEMYKILLNDYEVROKQIILMENAEKK 279

DB 1320 RQCAEAEEKFNRLRQOQER--LKTSKLSQDSLQTEQVNSLRDAKNVLENSLSEANARIEE 1377

QY 280 VLQOKKEMISLLSPQKKPRERAEDGTGTVAISDIEDDSGELSRDSVWGLSCDVTVR-EQ 338

DB 1378 -LQNAKV-----AQGNQLEAIRKLQEDAEKASRELQAKLEESTTSYES 1420

QY 339 LTNSTRKQWRILKSHVEKLDNQASKVHSEGLNEEDVISQDHEQETEKLEIEIERCKEMI 398

DB 1421 TINGLNEEITLKEELEKQIQOQLOQATSANEQNDLS---NIVSMKKSFEEDKIK-PI 1476

QY 399 KQAOQLLQOOLATTCDDTTSLRLDCYLLLEEKRLKEEWTL-FKQKQNFERRER-----451

DB 1477 KEKTOEVNEK-----ILEAQERLNQPSNIWNEBIKKWSEHEQEVSQ 1519

QY 452 -----RSFTEAAIRGLERKAFERASWVKQQLNMTNPDHQNSENKLFPS 498

DB 1520 KTRAEALKKRIPLTEKINKIRKKEELEK-----EFEEKVEERIKSME 1567

QY 499 AFGSGSD---PDNLIVHSPROKKLHSAVANGVPACTSKLTKSLPASPSPTS-----545

DB 1568 Q-SGEIDVVLKQLEAKVQEKQELENEYN-----KKLQEBELKDVPHSSHISDDERDKL 1620

QY 546 -----DPRQTHSCVSEHS-----SISVLNITPSESKPSEVARESTDQKWSQSRP 590

DB 1621 RAEIESRLREEFNQLAIKKKSFDEGKQOAMMKTTLLERKLAKMESQLSETKQSAESPP 1680

QY 591 SS 592

DB 1681 KS 1682

RESULT 10

T08621

centrosome associated protein CEP250 - human

C;Species: Homo sapiens (man)

C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C;Accession: T08621

R;Macq, G.J.; Rees, J.; Sandblom, O.; Balczon, R.; Fritszler, M.J.; Rattner, J.B.

A;Title: Autoantibodies to a group of centrosomal proteins in human autoimmune sera read

A;Reference number: Z16462; MUID:198165428; PMID:9506584

A;Accession: T08621

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-2442 <MAC>

A;Cross-references: UNIPARC:UPI0000071A1C; EMBL:AF022655; NID:g2832236; PIDN:AAC06349.1;

A;Experimental source: cell line HeLa

Query Match 7.0%; Score 220; DB 2; Length 2442;

Best Local Similarity 21.6%; Pred. No. 0.0068;

Matches 130; Conservative 103; Mismatches 181; Indels 188; Gaps 28;

QY 60 ENIQSISYLDQELTTFGFSPSLYEESKSKAKRELNIIVAVLNCMNEILLVQKKN-LLAQE 118

DB 1490 EHLPMVAOVERQKLTU-QREQIREPKDQRTQ--NVLE-----HOLLEKLEKQDMIESQ 1541

QY 119 SVETONLK-----LGSMDM---HLQSCYAKLKEQLETSRREM-----IGLQERD 159

DB 1542 RGQVODLKKQLVLTLECLALEENHHKWCQOKLIKELEGQRETQRTVALTHLTLEERS 1601

QY 160 RLQOKNRSRLHQL-----LKNKEDEVOKLQNIIASRATQYNHDKRKE 202

DB 1602 QELQAQSQIHDLSEHSVTLARELQERDQEVKSQREQIEELQ-----QKEHLTQDLERRD 1657

QY 203 REYNKLERHLQVLVMMKKDKNIAMDVLNVYGRADGKRGSWRTDKTEARNEDEMYKILLND 262

DB 1658 QELMLQKERIQ-----VLED 1672

QY 263 YEYRQKQIILMENAE-LKKVLQOKKEMISLLSPQKKPRERAEDGTGTVAISDIEDDSGE 321

DB 1673 QRTQTKLLEEDLEQIKLSLRGRG---LTTQRLMQERAEEGKGP---SKAQSGSLE 1725

QY 322 ----LSRDSVWGLSC-----DTVREQLTNSIRK--QMRILKSHVEK-----LDN 359

DB 1726 HMKLILRDKKEVECCQSHIHELQELQDLQQLQGLHRKVGETSLLSQREQEIVWLQQ 1785

QY 360 QASKVHSEG-LNEEDVTISRP-----HQEOTEKLELE-----IERCKEMIKAAQQ 403

DB 1786 QLQEAREGELGELSGLOSQDLQALQAOQDELEALQEQEQOQOAGQGBERVKEKADALQG 1845

QY 404 LLOQLATTCDDTTSLRLDCYLLLEKE---RLKEEWTL-----PKE 442

DB 1846 ALQNAHWTLKE-----RHGELQDHKEQARLEELAEVGRVQALBEVLGDLRAESRE 1898

QY 443 QKQNFERRERSFTAAIRLGLERKAFERASWVKQQLNMTNPDHQNSENKLFPSASG 502

DB 1899 QEKALLAQOQCAEQAOQAEHEVETRALQD---SWLQAQAVL-----KERDQLEALRAESQ 1950

QY 503 SSDPDNLIVHSPRPQ-----KKLHSAVANGVPAC---TSLKTKSLPASPSPT-----SDFRQT 550

DB 1951 SSRHQEEAARARAALQALQAKAHAAALQKGSQHLLEQAELSRSEASTATLQASLDACA 2010

QY 551 HS 552

DB 2011 HS 2012

RESULT 11

GB6266

hypothetical protein F3P19.25 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C;Accession: GB6266

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUD:21016719; PMID:11130712

A;Accession: GB6266

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1128 <STO>

A;Cross-references: UNIPROT:Q9SAF6; UNIPARC:UPI00000A7BC7; GB:AE005172; NID:g4850405; PI

C;Genetics:

A;Map position: 1

Query Match 6.9%; Score 218.5; DB 2; Length 1128;
Best Local Similarity 23.4%; Pred. No. 0.0032;
Matches 135; Conservative 108; Mismatches 196; Indels 139; Gaps 30;

QY 11 VLCTENKLSQYTSSETKMPSSLYSQVLCSSVPLSKNVHGVFGVFTGENI-----E 63
DB 105 LLLWENKEL--VSKHEQLNQAFQAEIL--KREQSHLYALTTVQREENLRKALGLEK 160
QY 64 QSIYSYLDQELTTGFPSPSLVSESK-----SKEAKR-ELN-IVAVLNCWN---ELLVLQRLK 114
DB 161 QCQVELEKALR-----EIQENSKIRLSSEAKLVEANALVASVNGRSSVDVENKIYSAESK 215
QY 115 LAQESVETQNLK-----GSDMDHLQSCYAKLKEQLE-TSRREMIGLQERDROLQCK 165
DB 216 LAEATRKSSELKURLEKVEVRESVLOQERLSFTKRESYEGTQKQREYLNWEKKLOQK 275
QY 166 NRSLLHOLLKN-----EK-DEVQKQLNIIASRATQYNHNV-----KKKEREYNKLEKRLHQ 214
DB 276 EESITEQKRLNQREBEKVNEIEKKLKEKELEEMNRKVDLSKSKSKETE-EDITKLEEE 334
QY 215 LVNKKDKNIAMDVLNVGRADGKRGSWT--DKTEARNEDEMYKILLNDYVROKQIL- 271
DB 335 LTTKEKEAHTLQITLL-----AKENELRAFBEKLIAREGTETQK-LIDD---QKEVLG 383
QY 272 -----MENAELKKVLQOMKXEMISLLSPQK-----KKPRERAEDGTGTVASIDIED 317
DB 384 SKMFELECEETIRKSLDKELQKIBELERQKVEIDHSEKLEKRNQ-----AMNKKFD 437
QY 318 DSGELSRDSVWGLSCDTVREQLTN-----SIRKQWRIL-KSHVEKLDNQASKVHSEGLN 370
DB 438 RYNKEMDLEAKLTKIKEREKIIQAEKELSLSEKQQLLSDKESLELDQOEIKIRAEWTK 497
QY 371 EEDVTSRQDHEQTEKLEIEIERCKEMIKAAQOQLQOQLATTCDDDTTSLRDCVLLERK 430
DB 498 KEEMI-----EBECKSLKEIKEREYRLQSELKSQIEKSRVHEEFLSKEVE-NLKQSK 551
QY 431 ERLKEEWTL-----FKEQKKNFER-----ERRSFTEAAR-----L 461
DB 552 ERFKEWEFLDKQAVYNKERTIRISEEKEKFERFOLLEGERLKKEBSALRVQIMQELDDI 611
QY 462 GLERKAFEEERASVWVKQQLNMTNFDHONS-----ENVKL 496
DB 612 RLQRESFE-----ANMEHRSALQEKVKL 635

RESULT 12
S74244
serine/threonine-specific protein kinase (EC 2.7.1.1-) isoform I, Rho-associated - mouse
C:Species: Mus musculus (house mouse)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 05-Oct-2004
C:Accession: S74244
R:Nakagawa, O.; Fujisawa, K.; Ishizaki, T.; Saito, Y.; Nakao, K.; Natumiya, S.
FEBS Lett. 392, 189-193, 1996
A:Title: ROCK-I and ROCK-II, two isoforms of Rho-associated coiled-coil forming protein
A:Reference number: S74244; MUID:96368048; PMID:8772201
A:Accession: S74244
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1354 <NAK>
A:Cross-references: UNIPROT:P70335; UNIPARC:UPI0000027020; EMBL:U58512; NID:g1514695; PI
F:74-338/Domain: coiled coil; phosphotransferase; serine/threonine-specific protein kinase
F:1229-1283/Domain: protein kinase C zinc-binding repeat homology <KIN>
F:1229-1283/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 6.9%; Score 218.5; DB 2; Length 1354;
Best Local Similarity 21.3%; Pred. No. 0.004;
Matches 147; Conservative 123; Mismatches 229; Indels 191; Gaps 33;

QY 61 NTEQISYLDQELTTGFPF-SUYESKSKAEKRLNIVAVLNCNELLVLQRLNLAQES 119
DB 476 NLESASVQIEKEMLLQHRINEYQKVEQENKRNINEVSTLKDLEDLRK---ASQT 532
QY 120 VETQNLKLGSDMDHLQSCYAKLKEQLETS---RREMIGLQERDROLQCKNRSLS---HQLL 173

Db 533 SOLANEKLTOLQKQEEANDLLRTESDTAVLRURKSHSTEMSKSISQLESNRELQERNRIL 592
QY 174 KNEKDEVQK---LQNIIASRATQYNHD-----VKKKEREYNKLEKRLHQL--- 215
DB 593 ENSKSOADKYQLOLVLEAERDRGHDSEMIGDLOARITSLOEEVVKHLKHNLERVEGER 652
QY 216 -----VMNKKDK---NIAMDVLNVGRADGKRGSWTDK---TEARNED----- 253
DB 653 KEAQMDLHSEKEKNLEID-LNYKLKSIQORLEQEVNEHKVTKARLTDKHQSIEBAKSV 711
QY 254 ---EMVKILLNDYEVYR-----OKQILMENAEKLVQOMKXEMISLLSPQKPKPR 300
DB 712 ANCEMEKKLKEBERAREKAENRVVETEKQCSMLDVLQK--SQQKLEHLT-----ENK 762
QY 301 ERAEDGTGTVASIDIEDDSGE-----LSRDSVWGLSCDTVRE----- 337
DB 763 ERMEDEVKNLAL-QLQESNKRLLQLLQNELKTOAFADNLKGLKQMKQBEINTLLEAKRLL 821
QY 338 -----QLTNSIRKQWR-----ILKSHVEKLDNQASKVHSEGL--- 369
DB 822 EFELAQLT-----KQVRNEGOMRELQDQLEAEQYFSTLYKTQVKELKEBIEEKNRENLRK 877
QY 370 -----NEEDVISRO-----DHEQETEKL-----ELEIERCKEMIKAAQOQLQOQLA 410
DB 878 IOELOSEKETLTQDLAETKAESQALGILLEEQYFELTQESKKAASNRQBITDKDHT 937
QY 411 TTCDDDTTSL--RDCVLL--EEKERL-----KEEWTLFKEO-----KKNFERERR 452
DB 938 VSRLETSNVLTKDTEMLARKENEELNEMRTAEVEYKLAKEEENLNKAAFEKNISTERT 997
QY 453 STEAAIRIG--LERKAFEEERASVWVKQQLNMTNFDHONSENVKLSAFSGSSDPDNL 510
DB 998 LKTOAVNKLAETMNRKDFKIDRKK-----ANTQDLRKKKEKNRKLQELNQEREXFNQM 1051
QY 511 VHSRPROKHLHSV-ANGUPACTSKLTSLPASPSTSDPFQTHSCVSEHSSISVLNITPPE 569
DB 1052 VVK--HQELNDMQALVEECTHRELQMLQASKESDIEQLRA-----KLDDLSDT 1101
QY 570 SKPSEVARETSQKWSQVSRPSSR--EGCVS 598
DB 1102 SVASFPSEADETD-----GNLPESRIEGWLS 1126

RESULT 13
T29095
cardiac muscle factor 1 - chicken
N:Alternate names: CMF1 protein
C:Species: Gallus gallus (chicken)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T29095
R:Wei, Y.; Bader, D.; Litvin, J.
Development 122, 2779-2789, 1996
A:Title: Identification of a novel cardiac-specific transcript critical for cardiac myo
A:Reference number: 220564; MUID:96379739; PMID:8787752
A:Accession: T29095
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1538 <WEI>
A:Cross-references: UNIPROT:Q98940; UNIPARC:UPI00000FC28E; EMBL:U62026; NID:g1621106; P
A:Experimental source: strain white leghorn; heart
C:Function:
A:Description: may be involved in cardiac myogenesis
C:Keywords: cardiac muscle; heart

Query Match 6.9%; Score 218.5; DB 2; Length 1538;
Best Local Similarity 21.8%; Pred. No. 0.0046;
Matches 154; Conservative 112; Mismatches 242; Indels 197; Gaps 36;

QY 15 ENKNLSQYTSKTPSSLYSQVLCSSVPLSK----- 47
DB 558 ENSGIKE-----KLESASVSKQQLSCGVVSLGKEPEKISAKHNVNKTVLSNTTDLDDIE 612

A/Accession: S34842
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 323-381, 'HKAI' <GUE2>
A/Cross-references: UNIPARC:UPI000017737D; EMBL:M28266
A/Note: difference at carboxyl end due to frameshift error
C/Comment: This protein is found as flocculent material in the parasitophorous vacuole.
C/Superfamily: trichohyalin; calmodulin repeat homology
C/Keywords: EF hand
F:154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-Q-S-D-L-E-Q-E-R-R)

Query Match 6.9%; Score 217.5; DB 2; Length 1909;
Best Local Similarity 24.7%; Pred. No. 0.0067;
Matches 125; Conservative 90; Mismatches 167; Indels 125; Gaps 25;

Qy 26 TMSPSLSYQQVLCSSVPLSKNVHGVFGVCTGENIEQISYLDQELTTTFGPSPSYERS 85
Db 1188 TKASKETLQEQSDLEQERLAK-----EKLQEQSDLEQ-----ER 1223

Qy 86 KSKEAKRELNIIVAVLNCMELLVLQRKNL-----LAQSVETQNLKLGSDMDHLQSCYAKL 141
Db 1224 RAKEKLEQ-----QQSDLEQERLAKKEKLEQEQ-----SDLEQERRAKEKL 1263

Qy 142 KEQ---LETSSREMIGLERDROLCKNRSLSHQL-----LKNKDEVOVKLQ----- 184
Db 1264 QEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQ 1323

Qy 185 -NIIASRATQYNHDKVRKEREYVKLKER---LHQLVMNKKDKNIAMDVLNYYVGRADGKRG 240
Db 1324 ERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKKEKLEQEQSDLEQERRAKEKLQ 1383

Qy 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAEKLVQMKKEMISLLSPQKKKPR 300
Db 1384 EQQSD---LEQDRLAKKEKLEQ---QQRDLQERRAKEKLQEQQSDL-----EQERRAK 1430

Qy 301 ERAEDGTGTVAISDIEDDSGELSRDSVWGLSCDTVREQLTNSIRKOWRILKSHVEKLDNQ 360
Db 1431 EKLQEQQ-----SDLEQER-----RAKEKLEQ---QSDLEQERRAK---EKLQEQ 1470

Qy 361 ASKVHSEGLNEEDVISRQ-DHEQE---TEKL-----ELEIE-RCKEMIKAQOQLLQOQILA 410
Db 1471 QSDLEQERLAKKEKLQEQQRDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQE-- 1528

Qy 411 TTCDDDDTTSLLRDCYLLEKERLKEWTLFKQKKNFERRS---FTEAAIRLGLERKA 467
Db 1529 RLANEKLEQQRD---LEQERRAKEK---LQEQSDLEQERRAKEKLQEQQSDLEQERRA 1582

Qy 468 FE--ERASWVKQOFLNMTNFDHONSE 492
Db 1583 KEKLEQQSDLEQERLAKKEKLQEQQRD 1609

Search completed: June 12, 2006, 19:15:11
Job time : 46 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 12, 2006, 19:05:44 ; Search time 301 Seconds
(without alignments)
1889.982 Million cell updates/sec

Title: US-10-644-084-2

Perfect score: 3165

Sequence: 1 MGDWTVTPVLCTENKNLS.....CYSGCSSAPSAHGRRDPLP 615

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 7.2.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3165	100.0	615	1 ADIP MOUSE	Q8vc66 mus musculu
2	3165	100.0	627	2 Q69ZV4 MOUSE	Q69zv4 mus musculu
3	2937	92.8	613	1 ADIP RAT	Q8cg22 rattus norv
4	2786.5	88.0	614	1 ADIP HUMAN	Q9y2d8 homo sapien
5	1558.5	49.2	545	2 Q6NRX3 XENLA	Q6nrk3 xenopus lae
6	1554	49.1	554	2 Q6NRK1 XENLA	Q6nrk1 xenopus lae
7	1355	42.8	576	2 Q6P942 BRARE	Q6p942 brachydanio
8	1148.5	36.3	495	2 Q4RWX4 TETNG	Q4rwx4 tetraodon n
9	1142	36.1	528	2 Q4T607 TETNG	Q4t607 tetraodon n
10	1096	34.6	480	2 Q4RU57 TETNG	Q4ru57 tetraodon n
11	362	11.4	398	2 Q6Z0V1 ORYSA	Q6z0v1 oryza sativ
12	343	10.8	382	2 Q8GW47 ARATH	Q8gw47 arabidopsis
13	338	10.7	373	2 Q5AE7 ARATH	Q5ae7 arabidopsis
14	300	9.5	276	2 Q9FIE0 ARATH	Q9fie0 arabidopsis
15	260	8.2	1813	2 Q513P3 ENTHI	Q513p3 entamoeba h
16	258	8.2	284	2 Q8GUK2 ARATH	Q8guk2 arabidopsis
17	251	7.9	1919	2 Q4RIP0 TETNG	Q4rip0 tetraodon n
18	243.5	7.7	1985	2 Q7RC59 PLAYO	Q7rc59 plasmodium
19	240.5	7.6	1370	2 Q73732 XENLA	Q73732 xenopus lae
20	239	7.6	762	2 Q5OR43 ENTHI	Q5or43 entamoeba h
21	238.5	7.5	1738	2 Q76329 DICDI	Q76329 dictyosteli
22	237	7.5	2332	2 Q22847 CAREEL	Q22847 caenorhabdi
23	236	7.5	1024	2 Q54L07 DICDI	Q54l07 dictyosteli
24	235	7.4	924	2 Q15738 DICDI	Q15738 dictyosteli
25	234.5	7.4	846	2 Q4V612 PLACH	Q4v612 plasmodium
26	234.5	7.4	1679	1 MLF2 YEAST	P40457 saccharomyc
27	233	7.4	987	2 Q8LI18 ORYSA	Q8li18 oryza sativ
28	232.5	7.3	709	2 Q9GRG1 TETTH	Q9grg1 tetrahymena
29	231	7.3	1980	2 Q6FWE0 CANGA	Q6fwe0 candida gla
30	229.5	7.3	1596	2 Q81J44 PLAF7	Q81j44 plasmodium
31	227	7.2	1738	2 Q51ED7 ENTHI	Q51ed7 entamoeba h

32	226.5	7.2	753	2 Q2SR10 MYCCA	Q2sr10 mycoplasma
33	226.5	7.2	798	2 Q5B8R8 EMENI	Q5b8r8 aspergillus
34	226.5	7.2	1175	2 Q8XNW6 CLOPE	Q8xnw6 clostridium
35	226	7.1	1354	1 ROCK1 RABIT	Q77819 o rho-aesoc
36	226	7.1	2612	2 Q815X5 PLAF7	Q815x5 plasmodium
37	224.5	7.1	1723	2 Q4N897 THEPA	Q4n897 theileria p
38	224	7.1	712	2 Q98SN5 CHICK	Q98sn5 gallus gall
39	224	7.1	800	2 Q96X03 EMENI	Q96x03 emericeella
40	224	7.1	2139	2 Q07569 ENTHI	Q07569 entamoeba h
41	223.5	7.1	3259	1 G0GB1 HUMAN	Q14789 homo sapien
42	223	7.0	1699	2 Q4UHB3 THEAN	Q4uhb3 theileria a
43	223	7.0	2042	2 Q6BUQ9 DEBHA	Q6buq9 debaryomyce
44	222.5	7.0	1558	2 Q4SEM9 TETNG	Q4sem9 tetraodon n
45	222	7.0	1946	2 Q97291 PLAF7	Q97291 plasmodium

ALIGNMENTS

RESULT 1

ADIP_MOUSE

ID ADIP_MOUSE STANDARD; PRT; 615 AA.

AC Q8VC66; Q8BG59; Q8C7X0; Q8K2F7;

DT 11-OCT-2004, integrated into UniProtKB/Swiss-Prot.

DT 01-MAR-2002, sequence version 1.

DT 07-FEB-2006, entry version 18.

DE Afadin- and alpha-actinin-binding protein (ADIP) (Afadin DIL domain-interacting protein).

GN Name=Sbx2ip;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muridae; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP NUCLEOTIDE SEQUENCE [MRNA], FUNCTION, SUBCELLULAR LOCATION,

RP INTERACTIONS WITH AFADIN AND ALPHA-ACTININ, AND TISSUE SPECIFICITY.

RC STRAIN=C57BL/6;

RX MEDLINE=22450629; PubMed=12446711; DOI=10.1074/jbc.M209832200;

RA Asada M., Irie K., Morimoto K., Yamada A., Ikeda W., Takeuchi M.,

RA Takai Y.,

RT "ADIP, a novel afadin- and alpha-actinin-binding protein localized at cell-cell adherens junctions.,"

RT J. Biol. Chem. 278:4103-4111(2003).

RN [2]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].

RC STRAIN=C57BL/6J; TISSUE=Brain cortex, Cerebellum, and Testis;

RC PubMed=16141072; DOI=10.1126/science.1112014;

RC Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,

RC Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,

RC Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,

RC Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,

RC Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,

RC Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,

RC Chiu K.L., Chowdhury V., Christoffels A., Clutterbuck D.R.,

RC Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,

RC di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,

RC Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,

RC Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,

RC Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,

RC Hill D., Humenicki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,

RC Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,

RC Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,

RC Kurochkin I.V., Larau L.F., Lazarevic D., Lipovich L., Liu J.,

RC Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,

RC Matsuura H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,

RC Mottagui-Farzi S., Mulder N., Nakano N., Nakai S., Nori F., Ohara O.,

RC Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,

RC Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,

RC Petrovsky N., Piazza S., Reed J.F., Reid J.F., Ring B.Z., Ringwald M.,

RC Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,

RA Schonbach C., Sekiguchi K., Sempke C.A., Seno S., Sessa L., Sheng Y.,

RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,

RA Spertling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,
RA Tamoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA Iida J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,
RA Tagami M., Waki K., Watanishi A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RL Science 309:1559-1563(2005).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA]
RC STRAIN=PV/N; TISSUE=Mammary gland, and Salivary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klauener R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko A., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -|- FUNCTION: Belongs to an adhesion system which plays a role in the
CC organization of homotypic, interneuronal and heterotypic cell-cell
CC adherens junctions (AJs). May connect the nectin-afadin and E-
CC cadherin-catenin system through alpha-actinin and may be involved
CC in organization of the actin cytoskeleton at AJs through afadin
CC and alpha-actinin.
CC -|- SUBUNIT: Interacts with afadin and alpha-actinin.
CC -|- SUBCELLULAR LOCATION: Localized at cell-cell adherens junctions
CC (AJs). Not found at cell-matrix AJs.
CC -|- TISSUE SPECIFICITY: Widely expressed.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL, AF532969; AA015015.1; -; mRNA.
DR EMBL, AK049080; BAC33536.1; -; mRNA.
DR EMBL, AK031356; BAC27363.1; -; mRNA.
DR EMBL, AK043865; BAC31684.1; -; mRNA.
DR EMBL, BC021749; AAH21749.1; -; mRNA.
DR EMBL, BC031527; AAH31527.1; -; mRNA.
DR Ensembl, ENSMUSG0000036825; Mus musculus.
DR MGI, MG1:2139150; Ssx2ip.
DR GO, GO:0005515; F:protein binding; IPI.
DR Cell adhesion; Coiled coil.
KW CHAIN 1 615 Afadin- and alpha-actinin-binding
FT protein.
FT /FTID=PRO_0000064456.
FT FT
FT COILED 126 227 Potential.
FT COILED 266 293 Potential.
FT COILED 375 461 Potential.
FT CONFLICT 47 47 K -> R (in Ref. 3; AAH31527).
FT CONFLICT 406 406 Missing (in Ref. 2).
FT CONFLICT 503 503 S -> SKPG (in Ref. 2; BAC33536).
FT SEQUENCE 615 AA; 70956 MW; CC41D707859191F CRC64;
SQ

Query Match 100.0%; Score 3165; DB 1; Length 615;
Best Local Similarity 100.0%; Pred. No. 5e-133;
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGDWMTVDPVLCATENKNLSQYTTSEKMPSSLYSQQVLCSSVPPLSKNVHGVFGVFCGCE 60
DB 1 MGDWMTVDPVLCATENKNLSQYTTSEKMPSSLYSQQVLCSSVPPLSKNVHGVFGVFCGCE 60
QY 61 NIEOSISYLDDELTTFGPPSLYEESKSEAKRELINIVAVLNCMELLVLQRKNLLAQSSV 120
DB 61 NIEOSISYLDDELTTFGPPSLYEESKSEAKRELINIVAVLNCMELLVLQRKNLLAQSSV 120
QY 121 ETQNLKLGSDMDHLQSCYAKLKEQLETSRRREMIGLQERDROLQCKNRS LHQLLNKDEKDEV 180
DB 121 ETQNLKLGSDMDHLQSCYAKLKEQLETSRRREMIGLQERDROLQCKNRS LHQLLNKDEKDEV 180
QY 181 OKLQNIITASRATQYNHVDVKKREYNKLERLHQLVMNKKNTAMDVLNYYGRADGKRG 240
DB 181 OKLQNIITASRATQYNHVDVKKREYNKLERLHQLVMNKKNTAMDVLNYYGRADGKRG 240
QY 241 SWRTDKTEARNEDEMYKILLNDYEVROKQILMENAEKLVLOQMKEMISLLSPQKKPR 300
DB 241 SWRTDKTEARNEDEMYKILLNDYEVROKQILMENAEKLVLOQMKEMISLLSPQKKPR 300
QY 301 ERAEDGTGTVAISDIEDSGELSRDSVWGLSCDVTREQLTNSIRKQWRLKSHVEKLDNQ 360
DB 301 ERAEDGTGTVAISDIEDSGELSRDSVWGLSCDVTREQLTNSIRKQWRLKSHVEKLDNQ 360
QY 361 ASKVHSEGLNEEDVISQDHPQTEKLELETERCKEMIKAOQQLLOQLATTCDDDTTSL 420
DB 361 ASKVHSEGLNEEDVISQDHPQTEKLELETERCKEMIKAOQQLLOQLATTCDDDTTSL 420
QY 421 LRDCVLLSEERLKEEWTLFKQCKNFERRRSFTFAAIRGLERKAFEEERASWVKQOF 480
DB 421 LRDCVLLSEERLKEEWTLFKQCKNFERRRSFTFAAIRGLERKAFEEERASWVKQOF 480
QY 481 LNMNTFQHNSENKVLFSAFSGSSDDPNLI VHSRPRQKLSHVANGVPACTSKLTSLPA 540
DB 481 LNMNTFQHNSENKVLFSAFSGSSDDPNLI VHSRPRQKLSHVANGVPACTSKLTSLPA 540
QY 541 SPSTSDFRQTHSCVSEHSSISVLNITPEESKPESEARESTQKWSVQSRPSREGCYSGC 600
DB 541 SPSTSDFRQTHSCVSEHSSISVLNITPEESKPESEARESTQKWSVQSRPSREGCYSGC 600
QY 601 SSAPRSAGHRDDLP 615
DB 601 SSAPRSAGHRDDLP 615
RESULT 2
Q69ZV4_MOUSE PRELIMINARY; PRT; 627 AA.
ID Q69ZV4_MOUSE
AC Q69ZV4;
DT 13-SEP-2004, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2004, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE MKIAA0923 protein (Fragment).
GN Name=Ssx2ip; Synonyms=mKIAA0923;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryonic tail;
RX PubMed=15368895; DOI=10.1093/dnares/11.3.205;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoaka S.,
RA Suga Y., Seino S., Nishimura M., Kaisho T., Hoshino K., Kitamura H.,
RA Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT IV. The complete nucleotide sequences of 500 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones

RT	randomly sampled from size-fractionated libraries.;									
RL	DNA Res. 11:205-218(2004).									
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms									
CC	Distributed under the Creative Commons Attribution-NoDerivs License									
CC	-----									
DR	EMBL; AK173064; BAD32342.1; -; mRNA.									
DR	Ensembl; ENSMUSG0000036825; Mus musculus.									
DR	MGI; MGI:2139150; Sex2ip.									
DR	GO; GO:0005515; F:protein binding; IPI.									
FT	NON_TER	1								
SQ	SEQUENCE	627 AA; 72053 MW; 7F2B9A2DF5F73E2D CRC64;								
Query Match										
Best Local Similarity 100.0%; Score 3165; DB 2; Length 627;										
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;										
QY	1	MGDMVTVPVLTCTENKNLSQYTSKMPSSLYSQVLCSSVPLSKNVHGVFGVCTGE	60							
DB	13	MGDMVTVPVLTCTENKNLSQYTSKMPSSLYSQVLCSSVPLSKNVHGVFGVCTGE	72							
QY	61	NIEQSIYLDQELTTFGFPLSYEESKSKAKRELNIIVAVLNCMNEELLVLQKNLLAQESV	120							
DB	73	NIEQSIYLDQELTTFGFPLSYEESKSKAKRELNIIVAVLNCMNEELLVLQKNLLAQESV	132							
QY	121	ETQNLKLGSDMDHLQSCYAKLKEQLETSRREMIGLQERDROCKRSLHQLLKNKDEV	180							
DB	133	ETQNLKLGSDMDHLQSCYAKLKEQLETSRREMIGLQERDROCKRSLHQLLKNKDEV	192							
QY	181	QKLQNIASRATQYNHDKVKKREYNKLERLHOLVMNKKDKNIAMDVNLVYGRADGKRG	240							
DB	193	QKLQNIASRATQYNHDKVKKREYNKLERLHOLVMNKKDKNIAMDVNLVYGRADGKRG	252							
QY	241	SWRTDKTEARNEDEMYKILLNDYEVYRQKQILMENAEKLVLOQMCKEMISLLSPQKKPR	300							
DB	253	SWRTDKTEARNEDEMYKILLNDYEVYRQKQILMENAEKLVLOQMCKEMISLLSPQKKPR	312							
QY	301	ERAEDGTGTVALSIEDDSELSRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ	360							
DB	313	ERAEDGTGTVALSIEDDSELSRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ	372							
QY	361	ASKVHSEGLNEEDVISRQDHEQTEKLEIEERCKEMIKAAQQLLQOQLATTCDDDTTSL	420							
DB	373	ASKVHSEGLNEEDVISRQDHEQTEKLEIEERCKEMIKAAQQLLQOQLATTCDDDTTSL	432							
QY	421	LRCVYLLEKRLKEEWTLFKQKKNFERERSFTEAAILRLGLERKAFEEERASWYKQF	480							
DB	433	LRCVYLLEKRLKEEWTLFKQKKNFERERSFTEAAILRLGLERKAFEEERASWYKQF	492							
QY	481	LNMTFDHQNSENVKLFSAFGSSDDPNLIVHSRPRQKKLHVSANGVPACTSKLTKSLPA	540							
DB	493	LNMTFDHQNSENVKLFSAFGSSDDPNLIVHSRPRQKKLHVSANGVPACTSKLTKSLPA	552							
QY	541	SPSTSDFRQTHSCVSHSHSISVLNITPEESKPESEVARESTDQKWSVQSPSPREGCYSC	600							
DB	553	SPSTSDFRQTHSCVSHSHSISVLNITPEESKPESEVARESTDQKWSVQSPSPREGCYSC	612							
QY	601	SSAFRSAHGDRDDL 615								
DB	613	SSAFRSAHGDRDDL 627								
RESULT 3										
ID	ADIP	RAT								
AC	Q8CG22;									
DT	11-OCT-2004,	integrated into UniProtKB/Swiss-Prot.								
DT	01-MAR-2003,	sequence version 1.								
DE	07-FEB-2006,	entry version 16.								
DE	Afadin- and alpha-actinin-binding protein (ADIP) (Afadin DIL domain-interacting protein).									
GN	Name=Sex2ip;									
OS	Rattus norvegicus (Rat).									
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Rattus.									
CC	NCBI_TaxID=10116;									
RN	[1]									
RP	NUCLEOTIDE SEQUENCE [MRNA]. FUNCTION, SUBCELLULAR LOCATION, AND INTERACTIONS WITH AFADIN AND ALPHA-ACTININ.									
RC	STRAIN=Sprague-Dawley;									
RX	MEDLINE=22450629; PubMed=12446711; DOI=10.1074/jbc.M209832200;									
RA	Asada M., Irie K., Morimoto K., Yamada A., Ikeda W., Takeuchi M., Takai Y.;									
RT	"ADIP, a novel afadin- and alpha-actinin-binding protein localized at cell-cell adherens junctions.";									
RL	J. Biol. Chem. 278:4103-4111(2003).									
RN	[2]									
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].									
RC	TISSUE=Testis;									
RG	NIH - Mammalian Gene Collection (MGC) project;									
RL	Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.									
CC	-!- FUNCTION: Belongs to an adhesion system, which plays a role in the organization of homotypic, interneuronal and heterotypic cell-cell adherens junctions (AJs). May connect the nectin-afadin and E-cadherin-catenin system through alpha-actinin and may be involved in organization of the actin cytoskeleton at AJs through afadin and alpha-actinin.									
CC	-!- SUBUNIT: Interacts with afadin and alpha-actinin.									
CC	-!- SUBCELLULAR LOCATION: Localized at cell-cell adherens junctions (AJs). Not found at cell-matrix AJs.									
CC	-----									
CC	Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms									
CC	Distributed under the Creative Commons Attribution-NoDerivs License									
CC	-----									
DR	EMBL; AF532970; AA015016.1; -; mRNA.									
DR	EMBL; BC078687; AAH78687.1; -; mRNA.									
DR	HSSP; P01096; 1HF9.									
DR	Ensembl; ENSRNOG00000015425; Rattus norvegicus.									
DR	RGD; 70849; Sex2ip.									
KW	Cell adhesion; Coiled coil.									
FT	CHAIN	1	613							
FT	Afadin- and alpha-actinin-binding protein.									
FT	/FTID=PRO_0000064457.									
FT	COILED	126	227							
FT	COILED	266	293							
FT	COILED	375	461							
SQ	SEQUENCE	613 AA; 70679 MW; 55A6174A810AB8DE CRC64;								
Query Match										
Best Local Similarity 92.8%; Score 2937; DB 1; Length 613;										
Matches 567; Conservative 18; Mismatches 28; Indels 2; Gaps 1;										
QY	1	MGDMVTVPVLTCTENKNLSQYTSKMPSSLYSQVLCSSVPLSKNVHGVFGVCTGE	60							
DB	1	MGDMVTVPVLTCTENKNLSQYTSKMPSSLYSQVLCSSVPLSKNVHGVFGVCTGE	60							
QY	61	NIEQSIYLDQELTTFGFPLSYEESKSKAKRELNIIVAVLNCMNEELLVLQKNLLAQESV	120							
DB	61	NIEQSIYLDQELTTFGFPLSYEESKSKAKRELNIIVAVLNCMNEELLVLQKNLLAQESV	120							
QY	121	ETQNLKLGSDMDHLQSCYAKLKEQLETSRREMIGLQERDROCKRSLHQLLKNKDEV	180							
DB	121	ETQNLKLGSDMDHLQSCYAKLKEQLETSRREMIGLQERDROCKRSLHQLLKNKDEV	180							
QY	181	QKLQNIASRATQYNHDKVKKREYNKLERLHOLVMNKKDKNIAMDVNLVYGRADGKRG	240							
DB	181	QKLQNIASRATQYNHDKVKKREYNKLERLHOLVMNKKDKNIAMDVNLVYGRADGKRG	240							
QY	241	SWRTDKTEARNEDEMYKILLNDYEVYRQKQILMENAEKLVLOQMCKEMISLLSPQKKPR	300							
DB	241	SWRTDKTEARNEDEMYKILLNDYEVYRQKQILMENAEKLVLOQMCKEMISLLSPQKKPR	300							
QY	301	ERAEDGTGTVALSIEDDSELSRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ	360							
DB	301	ERAEDGTGTVALSIEDDSELSRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ	360							

```
QY 361 ASKVHSEGLNEBDVISRODHEQTEKLEIEIRCKEMIKAOQLOQLATTCTDDTTSL 420
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 361 ASKVHSEGHEDVDVISRODHEQTEKLEIEIRCKEMIKAOQLOQLATTCTDDTTSL 420
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 421 LRDCYLLEKRLKEWTLFKQKNFRERRSFTEAAIRLGLERKAFEEERASVVKQOF 480
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 421 LRDCYLLEKRLKEWTLFKQKNFRERRSFTEAAIRLGLERKAFEEERASVVKQOF 480
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 481 LNMTHFDHONSENVKLFSAFSGSSDPDNLIVHSRPROKXKLVANGVPACTSKLTPSLPA 540
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 481 LNMTHFDHONSENVKLFSAFSGSSDPDNLIVHSRPROKXKLVANGVPACTSKLTPSLPT 540
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 541 SPSTDFRTHSCVSEHSHSVINLITPEESKPESEVARESTDQKWSVQRPSPSREGCYGSC 600
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 541 SP-SDFCPSRCVSEHSPVSAVTPTPEETKNEVGRESTDQKWSVQRPSPSREGCYGSC 598
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY 601 SNAFRSAHGRDRLP 615
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
DB 599 SSAYTSSHVERDLP 613
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
```

RESULT 4

```
ADIP HUMAN STANDARD; PRT; 614 AA.
ID ADIP HUMAN STANDARD; PRT; 614 AA.
AC Q9Y2D8; O6P2P8; O6ULS1; Q7L168; Q9UIX0;
DT 11-OCT-2004, integrated into UniProtKB/Swiss-Prot.
DT 11-OCT-2004, sequence version 3.
DE AFadin-2006, entry version 23.
DE Afadin- and alpha-actinin-binding protein (ADIP) (Afadin DIL domain-
interacting protein) (SSX2-interacting protein).
GN Name=SSX2IP; Synonyms=KIAA0923;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], INTERACTIONS WITH SSX2 AND SSX3, AND
RP SUBCELLULAR LOCATION.
RX PubMed=12007189; DOI=10.1002/gcc.10073;
RA de Bruijn D.R.H., dos Santos N.R., Kater-Baats E., Thijssen J.,
RA van den Berk L., Stap J., Balemans M., Schepens M., Merckx G.,
RA van Kessel A.G.;
RT "The cancer-related protein SSX2 interacts with the human homologue of
a Ras-like GTPase interactor, RAB3IP, and a novel nuclear protein,
RT SSX2IP."
RL Genes Chromosomes Cancer 34:285-298(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 1).
RC TISSUE=Testis;
RA Lu L., Huang X.Y., Xu M., Yin L.L., Li J.M., Zhou Z.M., Sha J.H.;
RT "Cloning a new transcript of X breakpoint 2 interacting protein
RT (SSX2IP) in testis."
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=99246063; PubMed=10231032; DOI=10.1093/dnares/6.1.63;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hiroseawa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:63-70(1999).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM 1).
RC TISSUE=Testis;
RG The German cDNA consortium;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORMS 1 AND 2).
```

```
RC TISSUE=Brain, and Eye;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Wozny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting J., Helton E., Kettelman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Fabry M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Belongs to an adhesion system, which plays a role in the
organization of homotypic, interneuronal and heterotypic cell-cell
adherens junctions (AJs). May connect the nectin-afadin and E-
cadherin-catenin system through alpha-actinin and may be involved
in organization of the actin cytoskeleton at AJs through afadin
and alpha-actinin (By similarity).
CC -!- SUBUNIT: Interacts with afadin and alpha-actinin (By similarity).
CC Interacts with SSX2 and SSX3. Does not interact with SSX1 and
SSX4.
CC -!- SUBCELLULAR LOCATION: Localized at cell-cell adherens junctions
(AJs). Not found at cell-matrix AJs (By similarity). Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9Y2D8-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9Y2D8-2; Sequence=VSP 011724, VSP 011725;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Widely expressed, with the highest expression
in brain, intermediate expression in kidney, testis, spinal cord,
liver, heart, lung, skeletal muscle, ovary, fetal liver and fetal
brain, and little to no expression in pancreas and spleen. All
specific brain regions showed intermediate to high expression,
with highest expression in amygdala. Also expressed in fetal
tissues, mainly in liver and brain.
CC -!- DOMAIN: Both the N-terminal (up to position 79) and the C-terminal
(from position 304) sequences are required for interaction with
SSX2.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY367055; AAQ72373.1; -; mRNA.
DR EMBL; AB023140; BAA76767.2; ALT_INIT; mRNA.
DR EMBL; AL133046; CAB61373.1; -; mRNA.
DR EMBL; BC033637; AAH33637.1; -; mRNA.
DR EMBL; BC064389; AAH64389.1; -; mRNA.
DR PIR; T42649; T42649.
DR Ensembl; ENSG00000117155; Homo sapiens.
DR HGNC; HGNC:16509; SSX2IP.
DR MIM; 608690; gene.
DR LinkHub; Q9Y2D8; -.
KW Alternative splicing; Cell adhesion; Coiled coil; Nuclear protein.
FT CHAIN 1 614 Afadin- and alpha-actinin-binding
FT protein.
FT /FTid=PRO_0000064455.
FT COILED 131 227 Potential.
FT COILED 266 293 Potential.
FT COILED 374 460 Potential.
FT VARSPPLIC 557 560
FT /FTid=VSP_011724.
```

```
FT VARSPLIC 561 614 Missing (in isoform 2).
FT CONFLICT 232 232 /FtId=VSP_011725.
FT CONFLICT 405 405 V -> A (in Ref. 2).
FT CONFLICT 576 576 Missing (in Ref. 5; AAH64389).
FT CONFLICT 578 578 G -> V (in Ref. 2).
FT CONFLICT 578 578 C -> R (in Ref. 4).
SQ SEQUENCE 614 AA; 71236 MW; 427903BF86A6FE31 CRC64;

Query Match 88.0%; Score 2786.5; DB 1; Length 614;
Best Local Similarity 87.5%; Pred. No. 3.5e-116;
Matches 538; Conservative 35; Mismatches 41; Indels 1; Gaps 1;

QY 1 MGDWMTVTDVPLCTENKLSOYSETKMSPSLYSQOVLCSYVPLSKNVHGVGVCCTGE 60
DB 1 MGDWMTVTDVPLCSSEKTSQYSETKMSPSLYSQOVLCSYVPLSKNVHGFSAFCTED 60

QY 61 NIEQISYLDQELTTFGFPSPLYEESKAEKRELINIVAVLNCNELLVLQKNLLAQESV 120
DB 61 NIEQISYLDQELTTFGFPSPLYEESKAEKRELINIVAVLNCNELLVLQKNLLAQENV 120

QY 121 ETQNLKLGSDMDHLQSCYAKLEQLETSRREMIGLQERDRQLQCKNRSIHLQKNEKDEV 180
DB 121 ETQNLKLGSDMDHLQSCYAKLEQLETSRREMIGLQERDRQLQCKNRSIHLQKNEKDEV 180

QY 181 OKLONIIASRATQYNDHVKREYNKLERHLQVNMKKDKNIAMDVLNYYGRADCKRG 240
DB 181 OKLONIIASRATQYNDHMKRKEYNKLERHLQVNMKKDKNIAMDVLNYYGRADCKRG 240

QY 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAEKLVQOMKEMISLLSPQKKPR 300
DB 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAEKLVQOMKEMISLLSPQKKPR 300

QY 301 ERAEDGTGTVAISDIEDSGELSRDSVWGLSCDTVREQLTNSIRKQWIRILKSHVEKLDNQ 360
DB 301 ERVDDSTGTV-ISDVEDAGELSRDSVWGLSCDTVREQLTNSIRKQWIRILKSHVEKLDNQ 360

QY 361 ASKVHSEGLNEEDVIRSDHEQTEKLELEIERCKEMIKAQOOLQOQLATTCDDDTTSL 420
DB 360 VSKVHLEGFNDEVDVIRSDHEQTEKLELEIOCKEMIKAQOOLQOQLATYDDDTTSL 420

QY 421 LRDCYLLEEKERLKEEWTLFKEQKNFERERSFTEAAILGLERKAFEEERASWVKQF 480
DB 420 LRDCYLLEEKERLKEEWSLFKEQKNFERERSFTEAAILGLERKAFEEERASWVKQF 479

QY 481 LNMTPDHQNSENVKLFSAFSGSDPDNLIIVHSRPRQKHLVANGVAPACTSKLTSLPA 540
DB 480 LNMTPDHQNSENVKLFSAFSGSDPDNLIIVHSRPRQKHLVANGVAPACTSKLTSLPA 539

QY 541 SPSTSDFRTHSCVSHSHSISVLNITPEESKPEVARESTDQKWSVQSPSSREGCYSGC 600
DB 540 SPSTSDFCQTRCISEHSHSINVLNITAEIKPNQVGECTNQKWSVASRPGSQEGCYSGC 599

QY 601 SSAFRSAHGDRLDLP 615
DB 600 SLSTYNHVEKDLP 614

RESULT 5
Q6NRX3 XENLA PRELIMINARY; PRT; 545 AA.
AC Q6NRX3_
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, sequence version 1.
DE MGC81138 protein.
GN Name=MGC81138;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE.
```

```
RC TISSUE=Ovary;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L.H., Shenmen C.F., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.M., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy M., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting J., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.C., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.W., Gough J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinaki M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Ovary;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391 (2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Ovary;
RA Klein S., Strausberg R.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC070587; AAH70587.1; -; mRNA.
SQ SEQUENCE 545 AA; 63836 MW; 79A2CB644DBA2CA3 CRC64;

Query Match 49.2%; Score 1558.5; DB 2; Length 545;
Best Local Similarity 55.7%; Pred. No. 1.3e-61;
Matches 326; Conservative 102; Mismatches 112; Indels 45; Gaps 13;

QY 1 MGDWMTVTDVPLCTENKLSOYSETKMSPSLYSQOVLCSYVPLSKNVHGVGVCCTGE 60
DB 1 MGDWMTISLP-----ESDKILQYSSBIRMSPTSLSS-----PSHSAANTLSGIVYNFCTED 52

QY 61 NIEQISYLDQELTTFGFPSPLYEESKAEKRELINIVAVLNCNELLVLQKNLLAQESV 120
DB 53 NIEQITVTDQELRTIGFTTQAVSKNGDG-RKLHLVSVINCIIHELLQNSQMSRKEEV 111

QY 121 ETQNLKLGSDMDHLQSCYAKLEQLETSRREMIGLQERDRQLQCKNRSIHLQKNEKDEV 180
DB 112 ETQLKINGDLEHLSIQORQKQOMEATRENCAQERDRQCKNRLQLLQKNEKEEV 171

QY 181 OKLONIIASRATQYNDHVKREYNKLERHLQVNMKKDKNIAMDVLNYYGRADCKRG 240
DB 172 OKLONIIASRSTQFNHVKRERENKLERLYQLVMDKDKISIDVLNYYGRADGKRT 231

QY 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAEKLVQOMKEMISLLSPQKKPR 300
DB 232 SWRTDKTAKNEENYKVLNDYEQQLMVENAEKLVQOMKEMISILS--QRKTK 289

QY 301 ERAEDGTGTVAISDIEDSGELSRDSVWGLSCDTVREQLTNSIRKQWIRILKSHVEKLDNQ 360
DB 290 EKLDSDIGPVA-SDIEEDLADSKENSELSCAEAREQLVSSIRQWRILKSHMEKLDNQ 348

QY 361 ASKVHSEGLNEEDVIRSDHEQTEKLETERCKEMIKAQOOLQOQLATTCDDDTTSL 420
DB 361 ASKVHSEGLNEEDVIRSDHEQTEKLETERCKEMIKAQOOLQOQLATTCDDDTTSL 420
```

Db 349 ATP-----DENGMIARGDHEQGLKLINEIQCKETIKIQOQLLKQFVSP-RDDTSTL 401
Qy 421 LDCYLLBEKERLKEBWTLFKQKKNFERERSFTEAATRLGLERKAFEEERASVVKQOF 480
Db 402 LQCYLLEDKERLQEBWKFNFQKKNFKERKNFTAAIRLGHKKAFFEDRAAWLKHQF 461
Qy 481 LNMNMF-DHONSENVKLFSAFSGSDPDNLI VHSRPROKHLHSHVANGVPACTSKLTSLP 539
Db 462 LNMVTFDHNKEE-KRANGVHFSPQDCHRLSHRTHDRHLASSGDHY-----QRKTLTP 516
Qy 540 ASPSTSDFRQTHSCVSEHSSISVLNITPES---KPSEVARESTD 581
Db 517 ITPS-----SKHS-----LTORESVAWRDSSISPNGTD 544

RESULT 6

Q6NRK1_XENLA
ID Q6NRK1_XENLA PRELIMINARY; PRT; 554 AA.
AC Q6NRK1;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, sequence version 1.
DE MGC83757 protein.
GN Name=MGC83757;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8335;
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Oocytes;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Schmutz J., Myers R.M.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN NUCLEOTIDE SEQUENCE.
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Oocytes;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative."
RL Dev. Dyn. 225:384-391(2002).
RN NUCLEOTIDE SEQUENCE.
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Oocytes;
RA Klein S., Strausberg R.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC070749; AAH0749.1; -; mRNA.
SQ SEQUENCE 554 AA; 64680 MW; 439EA3AEE3C2C98D CRC64;

Query Match 49.1%; Score 1554; DB 2; Length 554;
Best Local Similarity 55.2%; Pred. No. 2.2e-61;
Matches 321; Conservative 107; Mismatches 122; Indels 32; Gaps 10;
Qy 1 MGDMMVTDPVLCTENKNLSQVSTETKMSPSLSYQQVLCSSVPLSKNVHGVGVGFCTGEE 60
Db 1 MGDRTTSLP-----ESDKILOYSCEIRMSPTSLPS-----PSHVSANNLSGVSVYTFCTED 52
Qy 61 NIEQSISYLDQELTTFGPPSLYEESKSKAEKRELNIIVAVLNCMELLVLQRKNLLAQSRV 120
Db 53 NLEQCITYIDQELRTIGFPTQAVSKNGEG-RKLHLVSIINCIYELLQNSQTMRSNEEV 111
Qy 121 ETQNLKLGSDMDHLQSCYAKLKEOLETSRRRMIQLOERDROLQCKNRSIHLHLLKNEDEV 180
Db 112 ETQLKINGDLEYLOS IHORQDKQLEATKRENCALQERDROCKNRLQLLQKNEKEEV 171
Qy 181 OKLQNIIASRATQYNHDYKREYKRLKERLHQLVMNKKDKNIAMDVLNVYGRADGRG 240
Db 172 OKLQNIIASRSTQYNHSVKRERENYKRLERLYQLVMDKDKISIDVLNVYGRADGKRS 231
Qy 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLOOMKEMISLLSPQKKPR 300
Db 232 SWRTGKTDAKNEEMYKVLNDYEQKQKLVENVELKKVLOOMKEMISIVS--QRKTK 289
Qy 301 ERAEDGTGCTVAISDIEDDSGSLSDSVNGLSCDVTREOLTNSIRKOWRLKSHVEKLDNQ 360
Db 290 EKLEDSGTGTVT-SDIEEIIADSKENLSCEAVREQLISSIROQWRLKSHMEKLDNQ 348
Qy 361 ASKVHSEGLNEEDVISRDHEQETKLELEIERCKEMIKAAQQLLQQLQATTTCDTDTSL 420
Db 349 ACLNVPTPDENGLIAAEHEQELDKLISEIQCKETIIRSQQLLKQLSVPRDDDTSKL 408
Qy 421 LRDCYLLBEKERLKEBWTLFKQKKNFERERSFTEAATRLGLERKAFEEERASVVKQOF 480
Db 409 LQCYLLEDKERLQEBWKFNFQKKNFKERKNFTAAIRLGHKKAFFEDRAAWLKHQF 468
Qy 481 LNMNMF-DHONSENVKLFSAFSGSDPDNLI VHSRPROKHLHSHVANGVPACTSKLTSLP 539
Db 469 LNMVTFDHNKEE-KRANGVHFSPQDCHRLSHRTHDRHLASSGD-----YSRRPSKALP 523
Qy 540 ASPSTSDFRQTHSCVSEHSSISVLNITPESKPEVARESTD 581
Db 524 ITSS-----SKHSLTQIESIS---WRDSSISPNDDT 551
RESULT 7
Q6P942_BRARE
ID Q6P942_BRARE PRELIMINARY; PRT; 576 AA.
AC Q6P942;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Hypothetical protein zgc:73314.
GN ORFNames=zgc:73314;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN NUCLEOTIDE SEQUENCE.
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Wild-type; TISSUE=Eye;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN NUCLEOTIDE SEQUENCE.
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Oocytes;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative."
RL Dev. Dyn. 225:384-391(2002).
RN NUCLEOTIDE SEQUENCE.
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Oocytes;
RA Klein S., Strausberg R.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC070749; AAH0749.1; -; mRNA.
SQ SEQUENCE 554 AA; 64680 MW; 439EA3AEE3C2C98D CRC64;


```

QY 381 EOTEKLEIEIRCKEMIKAQOQLQOQLATTCDTDTSLRLDCCYLLLEKBLKEEWTLF 440
DB 355 EDEMREVOCKEFIOAQOQLQOQLNSLDDDTAFLHDCVTLLEKBLKEEWTLF 414
QY 441 KEQKNFERRRSFTAAIRLGLERKAFBEERASVWQOFLNMTNFDHQNSENVKLFSAF 500
DB 415 BEQKNFERKRNFTAAIRLGRD-----VRQSAL----- 444
QY 501 SGSSPDNLVHSRQPKLHVSANGVPACTSKLTKSLPASSTSDTFROTHSCVSEHSI 560
DB 445 -----PCARLIE-----CESLEBKGLPGGPSC-----LAESVSEHDSV 478

RESULT 9
Q4T607_TETNG
AC Q4T607_TETNG PRELIMINARY; PRT; 528 AA.
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 3.
DE Chromosome undetermined SCAF9008, whole genome shotgun sequence.
DE (fragment).
GN ORFNames=GSTENG0006544001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Parra G., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RL the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAEA01009008; CAF91675.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 528
SQ SEQUENCE 528 AA; 60582 MW; EA51CD5B7F4AD26 CRC64;

Query Match 36.1%; Score 1142; DB 2; Length 528;
Best Local Similarity 45.8%; Pred. No. 4.4e-43;
Matches 248; Conservative 96; Mismatches 111; Indels 86; Gaps 9;

QY 15 ENKNIQSQTSEKMPSSLSYQVLCSSVPLSKNVHGVFGVCTGENTEQ-----S 65
DB 4 EYKDVCGSPSECTPMPRFQO-----SSLPHQRTSYILSTFCTEHNVOECLLRISQITS 59
QY 66 ISYL--DQELTTTGFPSLYESKSKARELNIVAVLNCMNLVLRKNLLAQBSVETQ 123

```

```

DB 60 MSCLCVTCQAASGLPPVWPEPGSGS---EMNVAVLNCMDHLIOLYHRSLSRSLSELE 116
QY 124 NLKLGSDMDHLQSCYAKLKEOLETSRRRMIQLQERDROLOCKNRSLSHOLLKNEKEVQKL 183
DB 117 QHKSSSHVDYQLTSARLKEOLETSKRNTGLLRERRLQKAKSLQHSLSKNEKEVQKL 176
QY 184 QNIITASRATQYNDHVKRERENVKRLKRLHOLVMNKDKNIAMDVLNVVGRADGKRGSWR 243
DB 177 QNIITASRASQYNHMKRERENFKLKERLNQLLSDKREKKQAIQVNSIGRADGKRSLSWK 236
QY 244 TDKTEARNEDEMYKILLNDYEQKQILMENAELKVLQOMKEMISLLSPKKKPRERA 303
DB 237 TDKTEARHEGQLFKTLLSDYELRQRELLLENAELNKVLQOMKGMTSVLA--SNKSTLTG 294
QY 304 EDGTGTVAISDIEDSGELSRDSDVWGLSCDVTVRQOLTSIRKQWRLKSHVHEKLDNQASK 363
DB 295 DDGV-TQAKSEDEBEVDFSSKESV-ELFCVHAREKLTNSVRLQWKLNHVERLDSQAS- 351
QY 364 VHSEGLNEEDVISRQDHEQTEKLELETERCKEMIKAQOQLL----- 405
DB 352 -----LNNADVVSRETHEEVDKLTETIQOQFDLIQTOQQLLHVRLKLVQESSQGLGVN 406
QY 406 -----OOQLATTCDDDTTSLRLDCY 425
DB 407 HICHTCGICLLRCSDLIVCSEALDGAAGAAAANPFIVSQQQLSSQCDASSVHGSCP 466
QY 426 LLEEKERLKEEWTLFKEQKNFERRRSFTAAIRLGLERKAFBEERASVWQOFLNMTN 485
DB 467 MLQEKESLREEMKVLBEQKIFERRRNFTEAARLSDERKSFEEDRAWLKHQFLNLS 526
QY 486 F 486
DB 527 F 527

RESULT 10
Q4RU57_TETNG
ID Q4RU57_TETNG PRELIMINARY; PRT; 480 AA.
AC Q4RU57_TETNG
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Chromosome 1 SCAF14955, whole genome shotgun sequence. (fragment).
GN ORFNames=GSTENG00028943001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Parra G., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RL the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

```


SQ SEQUENCE 382 AA; 43509 MW; 9C7A54F700D515C7 CRC64;
Query Match 10.8%; Score 343; DB 2; Length 382;
Best Local Similarity 25.8%; Pred. No. 1.1e-07;
Matches 95; Conservative 92; Mismatches 139; Indels 42; Gaps 10;
QY 56 FCTGENIESISYLDQELTTGFPSPSYEESKSKAKRELINIVAVLNCNMLLVLRQKNLL 115
DB 28 FADVNLNENCIKYLNQSLVTSFSA----SLDLFATDPVSIARTCNVYALIQQORDVE 83
QY 116 AQSVEVTONKLGSDMDHLQSCYAKLKEQLETSRREMIGLQERDQLQCKNRSIHLQKN 175
DB 84 PRESANLQROQSDIARLEAKVERLEALLQCKOREIATITRTEAKNTAALKSQIEKLOQ 143
QY 176 EKDEVQKLQNIASR--ATQYNHVDVKRERYNKLERLHQLVM-NKDKQNIAMDVLNVY 232
DB 144 ERDEFQRM--VIANOQVKTQQLHETKKEKYIKLQERLNQVLMKEKXETSGMEIMNLL 201
QY 233 GRADGKRGSWRTDKTEARNEDEMYKILLNDYERQKQILMENAEELKVLQOMKKEMISLL 292
DB 202 QKEGRQRTGWSGKKTDS----DFYKKIVDAYEAKNQELMAENTDLRALLRSTQGMRSFL 257
QY 293 SPQKKKPRERABDG-----TGTVASIDIEDSGELSRDSVWGLSCDTVREQLTNSIR 344
DB 258 N-----ASGLTNOSLVANGRHGADPPSQSLG--GKTDVFDLPFRMARGQIEDSLR 306
QY 345 KQWRLKSHVEKLDNQASKVHSEGLNEEDVISRQDHEOTEKLELEIERCKEMIKAAQOOL 404
DB 307 SKQWSIKRGMGLVDAQEVS-----ITSEASERELE-LEAQLVEARSIIQEQSSI 356
QY 405 LQQQLATT 412
DB 357 MSKHLPKT 364

RESULT 13
Q5AE7 ARATH PRELIMINARY; PRT; 373 AA.
AC Q5AE7
DT 15-MAR-2005, integrated into UniProtKB/TrEMBL.
DT 15-MAR-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE At5G57410.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Kim C.J., Chen H., Cheuk R., Shinn P., Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (FEB-2005) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Kim C.J., Chen H., Cheuk R., Shinn P., Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (MAR-2005) to the EMBL/GenBank/DBSJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BT020622; AA081730.1; -; mRNA.
DR EMBL; BT021930; AA049379.1; -; mRNA.
SQ SEQUENCE 373 AA; 42659 MW; BE6A53693FB20F8 CRC64;
Query Match 10.7%; Score 338; DB 2; Length 373;
Best Local Similarity 24.7%; Pred. No. 1.8e-07;
Matches 89; Conservative 95; Mismatches 141; Indels 36; Gaps 9;
QY 56 FCTGENIESISYLDQELTTGFPSPSYEESKSKAKRELINIVAVLNCNMLLVLRQKNLL 115
DB 24 FANEDNLEHCTKYLNTQMTVTFGPPA----SLDLFSDNDPVSISRTCNMYSLLQQRORDIE 79

QY 116 AQSVEVTONKLGSDMDHLQSCYAKLKEQLETSRREMIGLQERDQLQCKNRSIHLQKN 175
DB 80 PRESANLQROQSDIARLEAKVERLEALLQCKOREIATITRTEAKNTAALKSQIEKLOQ 139
QY 176 EKDEVQKLQNIASR--ATQYNHVDVKRERYNKLERLHQLVM-NKDKQNIAMDVLNVY 232
DB 140 ERDEFQRM--VIGNQOVKAQOIHEMKKKEKDYIKLQERLNQVLMKEKXESRSGMEIMNLL 197
QY 233 GRADGKRGSWRTDKTEARNEDEMYKILLNDYERQKQILMENAEELKVLQOMKKEMISLL 292
DB 198 QKEGRQRTGWSGKKTDT----DFYKKIVDAYEAKNQELMAENTSLRALLRSTQGMRSFL 253
QY 293 SPQKKKPRERABDGTVASIDIEDSGELS-----RDSVWGLSCDTVREQLTNSIRKOWR 348
DB 254 N-----APNGSATLAGSEKREADPSQPLGKGTQVFDLPYRMARGQIESLRTKWA 304
QY 349 ILKSHVEKLDNQASKVHSEGLNEEDVISRQDHEOTEKLELEIERCKEMIKAAQOOLQO 408
DB 305 SIKESMVQLQAPKRAS-----VTSEATERELE-LEAQLVEARSIIQEQSISMSKH 354
QY 409 L 409
DB 355 L 355
RESULT 14
Q5F10 ARATH PRELIMINARY; PRT; 276 AA.
AC Q5F10
DT 01-MAR-2001, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2001, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MSF19.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99156233; PubMed=10048488; DOI=10.1093/dnares/5.6.379;
RA Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
RT Sequence features of the regions of 1,081,958 bp covered by seventeen
RT physically assigned P1 and TAC clones."
RL DNA Res. 5:379-391(1998).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB016891; BAB08478.1; -; Genomic DNA.
SQ SEQUENCE 276 AA; 31941 MW; 40104745551A5C5C CRC64;
Query Match 9.5%; Score 300; DB 2; Length 276;
Best Local Similarity 27.5%; Pred. No. 6.6e-06;
Matches 72; Conservative 71; Mismatches 97; Indels 22; Gaps 6;
QY 56 FCTGENIESISYLDQELTTGFPSPSYEESKSKAKRELINIVAVLNCNMLLVLRQKNLL 115
DB 12 FANEDNLEHCTKYLNTQMTVTFGPPA----SLDLFSDNDPVSISRTCNMYSLLQQRORDIE 67
QY 116 AQSVEVTONKLGSDMDHLQSCYAKLKEQLETSRREMIGLQERDQLQCKNRSIHLQKN 175
DB 68 PRESANLQROQSDIARLEAKVERLEALLQCKOREIATITRTEAKNTAALKSQIEKLOQ 127
QY 176 EKDEVQKLQNIASR--ATQYNHVDVKRERYNKLERLHQLVM-NKDKQNIAMDVLNVY 232
DB 128 ERDEFQRM--VIGNQOVKAQOIHEMKKKEKDYIKLQERLNQVLMKEKXESRSGMEIMNLL 185
QY 233 GRADGKRGSWRTDKTEARNEDEMYKILLNDYERQKQILMENAEELKVLQOMKKEMISLL 292

```
Db 186 QKEGRQGTWNGKKTDT----DFYKIVDAYEAKNQELMAENTSLRALLRSMQTDMDRDL 241
QY 293 SPOKKPRERAEDGTGTVAISD 314
Db 242 N-----APNGSATLAGSE 254

RESULT 15
ID Q513P3_ENTHI PRELIMINARY; PRT; 1813 AA.
AC Q513P3;
DT 07-JUN-2005, integrated into UniProtKB/TrEMBL.
DT 07-JUN-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Viral A-type inclusion protein repeat, putative.
GN ORFNames=82.t00025;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B.J., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Rongcaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Church C.M., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K.L., Ormond D., Squares R.,
RA Whitehead S., Quail M.A., Rabinowitsch E., Norbertczak H., Price C.,
RA Wang Z., Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S.,
RA Lohia A., Foster P.G., Sacheritz-Ponten T., Weber C., Singh U.,
RA Mukherjee C., El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M.,
RA Barrell B.G., Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAFB01000312; EAL48061.1; -; Genomic_DNA.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF02524; KID; 2.
KW Coiled coil.
SQ SEQUENCE 1813 AA; 214154 MW; 51D9CE79257621FB CRC64;

Query Match 8.2%; Score 260; DB 2; Length 1813;
Best Local Similarity 19.3%; Pred. No. 0.0029;
Matches 121; Conservative 114; Mismatches 188; Indels 204; Gaps 24;

QY 11 VLCTENKLSQYTSFTKMPSSLYSQVLCSYVPLSNVHGVGVCTGENTEQSISYLD 70
Db 452 IICDNNKEIAKFEQ-----ENLQKELNQIK 478

QY 71 Q-----ELTTGFPSPLYEESKSEAKREL--NIVAVLNCMVELLVORKNLLAOE 118
Db 479 EEKQKTENKVELVDVYTKQENELNKLKEEKEQIFNEKTTIENSINQ--IVBEKNKLTET 536

QY 119 -----SVETQNL-----KLGSMDHLSQCYAKLKEQLTSRR----- 150
Db 537 KESIKQELDSIKADNSTGKELEINKINEEKQLQNDYDTVQOEKENIQKELNQIKIEKSQK 596

QY 151 --EMIGLQERDRQLQCNRLHQLLKNKEDVOKLQNI-----IASRATQYNDV 198
Db 597 EEEELNKIKKEQOQVEDEKAKLITDIANGNDGLTKLNEVIDKLDEKENISNELNQIKNER 656

QY 199 KRKEREYNKLERLHQ-----LVNKKDKNIAMDVLNVGVR----- 234
Db 657 DNISNFPNKTKEIKQENETIQLNEEKSVLLNELNQIKKEEKQIEDEKAVIQOEKENEI 716
```

```
QY 235 -----ADKRGSWRTDKTEARN-----EDENVKIL-----LN 261
Db 717 TKLNEDKTVIENELNQIKTEKOEIENELNQIKDEKOKIEDEKSKLITELSNNGDGI SKLN 776
QY 262 DYEYRQKQILMENAEALKKVLQMKKEMISLLSPQKKKPRERADGTGTV-----AIS 313
Db 777 BELTQTKQ--EKENVNLNQIKNEFASFKEQNTQKENELKDNNKVOQELQKKNVEVS 833
QY 314 DIEDDSGELSRDSVWGLSCDTVRE-----QLTNSIRKOWRILKSHVEKLDNQASKV 364
Db 834 KLEEEKGNISNE----LS-NTKQELFQKKQEIITITQOEKEKENELKEQVKKIEEBSKL 888
QY 365 HSE-----GLAEEEDVISRODH-----EQETEKLELEIERCKEMIKAQOQLLOQQ 408
Db 889 ITELSNNGSDGISKLNEBELTQTKOEKEEIQKALEEEKEKLERIETELKEIKAEKQELVEEK 948
QY 409 LATTCDDDTTSLRDCYLLLEEKERLKEEWTLFKEQKKNFERRRRSFTEAAIRLGLERKAF 468
Db 949 NKTII--EETNLOQE---LNENKKIVEELTQTKOEKEEINNELNSIKE-----EKRI 996
QY 469 EEERASVWVKQFNLNMTNFDHNSENVK 495
Db 997 EEE-----KNQIINENK--EIKEENIK 1016
```

Search completed: June 12, 2006, 19:14:00
Job time : 306 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 12, 2006, 19:15:24 ; Search time 182 Seconds
(without alignments)
1565.259 Million cell updates/sec

Title: US-10-644-084-2
Perfect score: 3165
Sequence: 1 MGDWMTVTPVLTCTENKNLS.....CYSGCSSAFSAHGDRDDL 615

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
3: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
4: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
5: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
6: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2152.5	68.0	504	US-10-104-047-3467	Sequence 3467, Ap
2	2152.5	68.0	504	US-11-072-512-3467	Sequence 3467, Ap
3	379.5	12.0	443	US-10-425-114-54149	Sequence 54149, A
4	379.5	12.0	378	US-10-424-599-170119	Sequence 170119, A
5	377	11.9	389	US-10-425-115-341817	Sequence 341817, A
6	290	9.2	290	US-10-425-114-45937	Sequence 45937, A
7	243.5	7.7	1985	US-10-732-923-3351	Sequence 3351, Ap
8	238	7.5	2135	US-11-203-806A-12	Sequence 12, Appl
9	234.5	7.4	1679	US-10-369-493-22080	Sequence 22080, A
10	233	7.4	206	US-10-767-701-38884	Sequence 38884, A
11	233	7.4	987	US-10-437-963-180170	Sequence 180170, A
12	226.5	7.2	1175	US-10-732-923-3302	Sequence 3302, Ap
13	225	7.1	2503	US-10-828-985A-11	Sequence 11, Appl
14	222.5	7.0	3225	US-10-408-765A-254	Sequence 254, App
15	220.5	7.0	1875	US-10-369-493-22285	Sequence 22285, A
16	220.5	7.0	1875	US-10-732-923-3334	Sequence 3334, Ap
17	220.5	7.0	1875	US-10-732-923-3335	Sequence 3335, Ap
18	218.5	6.9	1128	US-10-732-923-3317	Sequence 3317, Ap
19	218	6.9	2383	US-10-082-830-3260	Sequence 3260, App
20	217.5	6.9	689	US-10-108-605-305	Sequence 305, App
21	217.5	6.9	744	US-11-097-143-40755	Sequence 40755, A
22	217.5	6.9	746	US-10-732-923-3339	Sequence 3339, Ap
23	217.5	6.9	1379	US-10-205-219-5	Sequence 5, Appl1
24	217.5	6.9	1909	US-10-732-923-3341	Sequence 3341, Ap
25	216.5	6.8	1132	US-10-732-923-3315	Sequence 3315, Ap
26	216.5	6.8	2665	US-11-124-368A-214	Sequence 214, App
27	216.5	6.8	2668	US-11-124-368A-215	Sequence 215, App

ALIGNMENTS

RESULT 1

US-10-104-047-3467
; Sequence 3467, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3467
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3467

Query Match 68.0%; Score 2152.5; DB 4; Length 504;
Best Local Similarity 90.3%; Pred. No. 3.6e-132;
Matches 419; Conservative 20; Mismatches 20; Indels 5; Gaps 2;
QY 1 MGDWMTVTPVLTCTENKNLSOYTSKTPSSLYSQOVLCSVPISKNVHGVGVCCTGE 60
DB 1 MGDWMTVTPD-----ESKTISQYTSKTPSSLYSQOVLCSVPISKNVHVSFFSFCED 56
QY 61 NIEQISYLDQELTTFGPPSLYEESKSEAKRELNI VAVINCWNEILLVLRKNLLAQESV 120
DB 57 NIEQISYLDQELTTFGPPSLYEESKSEAKRELNI VAVINCWNEILLVLRKNLLAQENV 116
QY 121 ETQNLKSGMDHLQSCYAKLKEQLETSRREMIGLOERDQLQCKNRSRLHQLKNEKDRV 180
DB 117 ETQNLKPGSDHLQSCYKLEQLETSRREMIGLOERDQLQCKNRLHQLKNEKDEV 176
QY 181 OKLQNIASRATQYVNDKREYKLERHQLVMNKDKNIAMDVLNVYGRADKRG 240
DB 177 OKLQNIASRATQYVNDKREYKLERHQLVMNKDKNIAMDVLNVYGRADKRG 236
QY 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLQOMKEMISLLSPQKKPR 300
DB 237 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLQOMKEMISLLSPQKKPR 296
QY 301 ERAEDGTGTVASIDEDSGELSRDSVWGLSCDTVREQLTNSIRKOWRILKSHVEKLDNQ 360
DB 297 ERVDDSTGTV-ISDVEEDAGELSRDSVWGLSCDTVREQLTNSIRKOWRILKSHVEKLDNQ 355
QY 361 ASKVHSEGLNEEDVLISRODHEQETKLELETERCKEMIKAOQOILLQOQLATTCDDDTSL 420

Db 356 VSKVHLEGFNDEVISRODHEQTEKLELEIQCKEMIKTQOQLQQQLATAYDDTTSL 415
QY 421 LRDCCYLLBEKERLKEEWTLFKQCKKNFERRRSFTEAAIRLGLE 464
Db 416 LRDCCYLLBEKERLKEEWSLFKQCKKNFERRRSFTEAAIRLGLE 459

RESULT 2

US-11-072-512-3467

; Sequence 3467, Application US/11072512

; Publication No. US20060029945A1

; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO

; APPLICANT: SUGIYAMA, TOMOYASU

; APPLICANT: OTSUKI, TETSUJI

; APPLICANT: WAKAMATSU, AI

; APPLICANT: SATO, HIROYUKI

; APPLICANT: ISHII, SHIZUKO

; APPLICANT: YAMAMOTO, JUN-ICHI

; APPLICANT: ISONO, YUUKO

; APPLICANT: HIO, YURI

; APPLICANT: OTSUKA, KAORU

; APPLICANT: NAGAI, KEIICHI

; APPLICANT: IRIE, RYOTARO

; APPLICANT: TAMECHIKA, ICHIRO

; APPLICANT: SEKI, NAOHITO

; APPLICANT: YOSHIKAWA, TSUTOMU

; APPLICANT: OTSUKA, MOTOTYUKI

; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO

; TITLE OF INVENTION: Novel full length cDNA

; FILE REFERENCE: 084335-0191

; CURRENT APPLICATION NUMBER: US/11/072,512

; CURRENT FILING DATE: 2003-03-07

; PRIOR APPLICATION NUMBER: US 60/350,978

; PRIOR FILING DATE: 2002-01-25

; PRIOR APPLICATION NUMBER: JP 2001-379298

; PRIOR FILING DATE: 2001-11-05

; NUMBER OF SEQ ID NOS: 4096

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3467

; LENGTH: 504

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-11-072-512-3467

Query Match 68.0%; Score 2152.5; DB 6; Length 504;

Best Local Similarity 90.3%; Pred. No. 3.6e-132;

Matches 419; Conservative 20; Mismatches 20; Indels 5; Gaps 2;

QY 1 MGDWMTVTPVLCTENKNLSOYTSKSPSSLYSQOVLCSVPKSNVHGVGVCCTGE 60
Db 1 MGDWMTVTPD-----BSKTSQYTSKSPSSLYSQOVLCSVPKSNVHGSFSAFCTED 56
QY 61 NIEQISYLDQLTTFGPPSLYEESKSEAKRELINAVLNCMNEILLVLRKNLLAQESV 120
Db 57 NIEQISYLDQLTTFGPPSLYEESKSEAKRELINAVLNCMNEILLVLRKNLLAQENV 116
QY 121 ETQNLKGSMDHLOSCYAKLKEQLETSSREMIGLOERDQLOCKNRSLHQLLKXNKEDEV 180
Db 117 ETQNLKGSMDHLOSCYAKLKEQLETSSREMIGLOERDQLOCKNRSLHQLLKXNKEDEV 176
QY 181 OKLQNIASRATQYVNDHVKREYNKLERHQLVMNKDKNIAMDVLNYVGRADGKRG 240
Db 177 OKLQNIASRATQYVNDHVKREYNKLERHQLVMNKDKNIAMDVLNYVGRADGKRG 236
QY 241 SWRTDKTEARNEDEMYKILLNDYERQKQILMENAEKLVQCKKEMISLLSPKKKPR 300
Db 237 SWRTCKTEARNEDEMYKILLNDYERQKQILMENAEKLVQCKKEMISLLSPKKKPR 296
QY 301 ERAEDGTGTVAISDIEDSGELSRDSVWGLSCDTVREQLTNSIRKQWIRLKSHEKLDNQ 360
Db 297 ERVDDSTGTV-ISDVEEDAGELSRDSMDLSCTVREQLTNSIRKQWIRLKSHEKLDNQ 355

QY 361 ASKVHSEGLNEEDVLSRODHEQTEKLELEIERCKEMIKAQOQLLQQQLATTCCDDTTSL 420
Db 356 VSKVHLEGFNDEVISRODHEQTEKLELEIQCKEMIKTQOQLQQQLATAYDDTTSL 415
QY 421 LRDCCYLLBEKERLKEEWTLFKQCKKNFERRRSFTEAAIRLGLE 464
Db 416 LRDCCYLLBEKERLKEEWSLFKQCKKNFERRRSFTEAAIRLGLE 459

RESULT 3

US-10-425-114-54149

; Sequence 54149, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 54149

; LENGTH: 443

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3150-041-D8_FLI pep

; US-10-425-114-54149

Query Match 12.0%; Score 379.5; DB 4; Length 443;

Best Local Similarity 26.6%; Pred. No. 2e-16;

Matches 106; Conservative 101; Mismatches 135; Indels 57; Gaps 14;

QY 27 KMSPSSLY-----SQOVLCSVPKSNVHGVGVCCTGENTIEQISYISLDQLTTFGFFSL 81
Db 54 RMSSSARFDLRASSQOL--QPPPSHASMSDGGAFANAENLEHCARYLNQTLVTFGPPA- 110
QY 82 YEEKSKSEAKRELINAVLNCMNEILLVLRKNLLAQESVETQNLKLGSDMDHLOSCYAKL 141
Db 111 ---SLDLFATDPVSIARTCNCIYALLOQRDIEFRSTNDQRQRMQSDISRLEAKIERM 167
QY 142 KEQLETSSREMIGLOERDQLOCKNRSLHQLLKXNKEDEVQKQNIASR--ATQYNHDKV 199
Db 168 DAQLAAKDRELATLTRTEAKNTAALKSQIDKLOQERDEFQKM--VIGNQVVRTQOIHEMK 225
QY 200 RKEREYNKLERLHQLVMNKDKN--TAMDVLNYVGRADGKRGSRWTDKTEARNEDEMYK 257
Db 226 KKEKEYIKLOPKLQVLMKEKKSSRSGMETMNLQKEGRQGRGTWNGKK---NDNDYYK 281
QY 258 ILLNDYERQKQILMENAEKLVQCKKEMISLLSPKKKPREAEDGTGTVAISDI-E 316
Db 282 MIVDAYEVKKQELMQENADRLRLSRMQMDNRDFN-----APNGSSQSTVTDNGR 332
QY 317 DDSGEL-----SRDSVWGLSCDTVREQLTNSIRKQWIRLKSHEKLDN--QASKVHSEGL 369
Db 333 QESSGPSQSLGGKTDVFDLPFPMARDQIEESLRTKMTSIKARMTQLQDAQKGAETVSEAT 392
QY 370 NEEDVISRODHEQTEKLELEIERCKEMIKAQOQLLQQQ 408
Db 393 DRE-----LELE-----AQLVEA-RSIIQEQ 412

RESULT 4

US-10-424-599-170119

; Sequence 170119, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

```

; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 170119
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_124632C.1.pep
US-10-424-599-170119

```

Query Match	12.0%	Score 379;	DB 4;	Length 378;
Best Local Similarity	26.3%	Pred. NO. 1.8e-16;		
Matches 94; Conservative	95;	Mismatches 143;	Indels 26;	Gaps 8;

Qy	56	FTCGENIEOSISYLBDEUUTTGFPSPSLVEESKSKAEKRELNIIVAVLNMOMELLVQLQRKNLL	115
Dd	25	FADADNLEHCAYLNQSLWTEGFFA---SLDLPANDPVSIIARTCNCIYVLLQORQORDE	80
Qy	116	AQESVETQNILKGSDMHLQSCYAKLKEOLESSTREMIQEDRDRQLQCKNRSLHOLLKN	175
Dd	81	PRESANDQORILLSDISLEAKVERLEGOLQVKDREIATITRTAEKNTAAKQAQIEKLQQ	140
Qy	176	EKDEVOKLQNIITASB--ATOYNHDVKKEREYNKLKERLHOLVM-NKKDKNTAMDVLNVV	232
Dd	141	ERDEFQRM--VIGNQOVQTQOMHEWKKEEYIKLQERLNOVLMEKKGSRSNGEMTNLL	198
Qy	233	GRADGKRGSWRDTKTARENEDEMYKILINDYEYRQKQILMENAEKLVQMCKEMISLL	292
Dd	199	QKEGRQRGTWGKA---DNDFYKXIVDAVESKNQELMAENADLRALLERSMQVDMDREL	254
Qy	293	SPOKKKPRERAEDGTGTVAISDIEDDSGELSRDSRWGLSODTVREQLTNSIRKQWRILKS	352
Dd	255	NAPNGLPK--QSFTGNERVESDPQSPLVGKMDFDLPPHMARDQIEESLRNKMASIKE	311
Qy	353	HVEKLDNQASKVHSEGLNEEDVISRQDHEQETEKLELEIERCEKMTAKAQOOLLQOOLA	410
Dd	312	RMVOLF-----QDAQEAEVTSATEARELE-LEAQLVARSIIQEASINSMKILA	359

```

RESULT 5
US-10-425-115-341817
; Sequence 341817, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 341817
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_74904C.1.pep
US-10-425-115-341817

```

Query Match 11.9%; Score 377; DB 4; Length 389;
Best Local Similarity 27.0%; Pred. NO. 2.5e-16;
Matches 99; Conservative 95; Mismatches 123; Indels 50; Gaps 12;

```

Qy 54 GVFTGENIEOSISVYLDOELTTTFPPSPSYESKSKEAKRELNAVVLNCMELLVLORKN 113
Db 30 GAFANAENLEHCARYLNTLVTFGPPA-----SLDLFATDPVSIAIARTCNCIIYALLQOQRD 85
Qy 114 LLAQESVETONLKLGSDMDHLOSACVAKLKEOLETSRREMIQLOERDRLOQCKNRSLOLL 173
Db 86 IEFRESTNDQORWOSDISRLEAKTIERMDQAUKADRELATUTRYEAKNTAAKLSQIDKL 145
Qy 174 KNEDEVOQLONIITASR--ATQYNHDVKRKEREYKNLKERLHOLVNKKDKN--IAMDVLT 229
Db 146 QOEDEFQKM--VICNQOVRTQQIHEMKKKEKEYIKLOEKUNOVLMEKKESSRSQMEIM 203
Qy 230 NYVGRAOGKGSWTDTKEARNEBEMYKILINDVEYROQILMENAEKKVILQOMKXEM 289
Db 204 NLLQEGRQRTWGNGK---NNDNYYYKMIVDAYEVKKQELMOENADLRALLRSQMOMDR 259
Qy 290 SLLSPQKKPRAEDGTGTVAISDI-EDDSGEL-----SRDSVNLSCDTVREQLTNSI 343
Db 260 DFLN-----APNGSSQTVTDNGHQSGSQSPOSPLGGKTVDVFDLPFHMMARDQIESL 310
Qy 344 RKOWRLKSHVEKLDN--QASKVHSEGLNEDVISRQDHQEOTEKLEIERCKEMIKAQ 401
Db 311 RTKMTSİKARTQLODAQKGAVTSEATDRE-----LELE----AQLVEA- 351
Qy 402 QOLLQOQ 408
Db 352 RSIIQEO 358

```

```

RESULT 6
US-10-425-114-45937
; Sequence 45937, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 45937
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700894116_FLI.pep
US-10-425-114-45937

```

Query Match	9.2%	Score 290;	DB 4;	Length 290;
Best Local Similarity	25.0%;	Pred. No. 8.2e-11;		
Matches	72;	Conservative 82;	Mismatches 112;	Indels 22; Gaps 7;
Qy	126	KLGSMDHLQSCYAKLKEQLQETSRREMIGLQERDRQLQCKNRSLHQLLKNKDEQVQLQN	185	
Db	3	RLLSDISRLCAKVERLEGQLQVDRBIATITRTEAKNTAAKQAEIKLQERDEFQRM--	60	
Qy	186	LIASR--ATQYNHDVKRKERYNKLERLHOLYV-NKKDKNIAMDVLNVYGRADGREGSW	242	
Db	61	VIGNQVKVTOQHEMKKKEKYIKLQERLNOVLMEKKESRSQWEINLLQEKGRQKGTW	120	
Qy	243	RTDKTEARNEDENWKILLNDYVYRQKOLIMENALKKVLQOMKEMISLLSPQKKPRPR	302	
Db	121	NGKKA---DNDFYKKIVDAYESKNQELMAENADLRALLRSQVDRDFLNAFGLPK--	174	
Qy	303	AEDGTGTVAISDIEDDSGELSRDSVWGSLSCDTVREQLTNSIRKQWRILKSHVFKLNDQAS	362	

Db 175 -OSFTGNVERVSPDSQSLVGRMDVFDLPFHMDQIBESLNRKVASIKERVMVL----- 228
Qy 363 KVHSEGLNEEDVLSRODHEQTEKLETERCKEMIKAOQQLLOOLA 410
Db 229 ----QDAQKAEVTSATERELE-LEAQLVEARSIIQEQASIMSKHLA 271

RESULT 7

US-10-732-923-3351
; Sequence 3351, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3351
; LENGTH: 1985
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
US-10-732-923-3351

Query Match 7.7%; Score 243.5; DB 5; Length 1985;
Best Local Similarity 19.3%; Pred. No. 1.1e-06;
Matches 110; Conservative 118; Mismatches 180; Indels 161; Gaps 21;
Qy 12 LCTENKNLSQYTSKMSPLSYSSQVLGSSVPLSKNVHGVFGVCTGTENIPQSISYL-D 70
Db 453 VCOASKQMARLQHIIESRQMLETEKML-----LKKK-----KENPEKVGMLAD 497
Qy 71 QELTFGFPSPSYEE-----SKSKAKRELINIVAVLNCWELL 107
Db 498 KEKEIDKIHQISQIKESLSINKNNETDEKEYVNSIKSYDNAQEL-----LDQNECI 552
Qy 108 VL-----ORKNLLAQSVEFQNLKLGSDMDH 133
Db 553 LIENCKSKLYDEYKFGQFNKKIIEIEREKEIEOERKNIIEKGNMLNENRR-----EIDE 609
Qy 134 LQSYAKLKEQLETSRREMIGLO-ERDQLOCKNRL-----HQLLNKEDEVQKLQNTIA 188
Db 610 EKLMMNKEKNELEMLKKELESLEKKEKKIIDCEYNNLQNKBELRNERNLLIK-ENELK 668
Qy 189 SPATQYNHVDKREKREYNKLERHLQVNNKQD-KNIAMDVLNYYGRADCKGSGWRTDKT 247
Db 669 NRIDKYN-----ELIDELNKNKKEIENDKMKMLNDI----- 699
Qy 248 EARNDEMYKILLNDYERQKQILMENAEKVKVLOQMKEMISLLSPKKPRERAED-- 305
Db 700 ----QDERIK-LJNETNNIKKE--NEKEINYKKEIKERISMINDEVKMKRLMEDIE 751
Qy 306 GTGTVAISDIEDSGELSRDSVWGLSCDTVRBQLTNSIRKQWRILKSHVFKLDNOASKVH 365
Db 752 NTKNIMLEDMEXNTK-----IKKEIENDKMMIKNIEDEKEYKTYLEESKF 798
Qy 366 SGLENEEDVISQDHEQTEKLELEIERCKEMI-----KAQQLLOQLATTCCDDTT 418
Db 799 NENLEKESELOQKYDDENNRLOAEINNEKKINKERNLNEKQKQKYVEDEFNCKEYEE 858
Qy 419 SLRDICYLLEKE-----RLKEEWTLFKBQKKNF-----BRERSFTEAARLGLERKA 467
Db 859 DRKKNYMLEEENNNKTRIMKEQELENYKKNVLDIEEEKDKLYVQOEKINLENILL 918
Qy 468 FEEERASWYKQFLNMTNFDHNSNVKL 496
Db 919 VEKEQ---IDIELKQFNKFNKEEENDIKI 944

RESULT 8

US-11-203-806A-12

; Sequence 12, Application US/11203806A
; Publication No. US20060051737A1
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Silverman, Joshua
; TITLE OF INVENTION: Nucleic and Amino Acid Sequences of
; TITLE OF INVENTION: Human Rf1 and Methods of Use Thereof
; FILE REFERENCE: 600-1-298N
; CURRENT APPLICATION NUMBER: US/11/203,806A
; CURRENT FILING DATE: 2005-08-15
; PRIOR APPLICATION NUMBER: 60/601,405
; PRIOR FILING DATE: 2004-08-13
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 2135
; TYPE: PRT
; ORGANISM: Fugu rubripes
US-11-203-806A-12

Query Match 7.5%; Score 238; DB 6; Length 2135;
Best Local Similarity 21.2%; Pred. No. 2.7e-06;
Matches 125; Conservative 122; Mismatches 228; Indels 116; Gaps 24;
Qy 86 KSEAKRELINIVAVLNCMELLVLRKNLLAQESVETQNLKLGSDMDHLQSCYAKLKEQL 145
Db 1522 QAQEAARKCGIAE--QKLNLDILSKNRQDILAE-----KLKEEFENAKKCAQEAQAK 1572
Qy 146 ETSRREMIGLO-----ERDQLOCKNRLHQLLNKED-EVQKLQNTIASRA 191
Db 1573 EAABEREAASLRQKAEAEKQKAAAEAAKQAKAEMLKKEAEVEVSKR---TAAQA 1629
Qy 192 TOYNHD-----VKREREYNKLER-----LHOLVNNKKDKNIAMDVLN-YVGRA 235
Db 1630 TALQKQKQADDEMAKQKQAEALQKSLVEKELTVVKLQLEKTEKO--MDVLDELQRV 1687
Qy 236 DGRGSRWRTDKTEARNEDEMYKILLNDYERQKQILMENAEKVKVLOQMKEMISLLSPQ 295
Db 1688 KGEVND--AIKQKQAEELSKVKTENSELLKLKIEE-ENRHLQKDKDKMQQLAE 1744
Qy 296 KKKPRERAEDGTVAISDIEDSGELSRDSVWGLSCDTVRBQLTNSIRKQWRILKSHVE 355
Db 1745 AAKMKLLAEDA-----ARLSVEATELARQMAESDLAEQRAJSEKMLKEMQVSEAT 1798
Qy 356 KLDNOASKVHSEGLNEEDVISR--QDHEQTEKLELEIERCKEMIKAOQQLLOQLATTTC 413
Db 1799 KLKAEAELOQRNOAQEMANKLODKQKIQERLDQETEAQKSLAEQ---KRQLEISA 1855
Qy 414 DDDTTSL-----LRDCYLLEKE-----RLKEEWTLFKBQKKNF-----E 448
Db 1856 ETEALKLVKVELTDSNAKAEDEVKVKRQSDDEVVKVQLQAEIKQNKIEVLQKSETQTLQSS 1915
Qy 449 RERSFTEAARLGLERKAPFEERASVWQK-----QFLNMTNFDHNSNVKLFSAF 500
Db 1916 REAESLRKAVADLBEKEREQLKKEAELOKQSPSKRLDLEALMGPPDVNHSPPSHVR--GTW 1973
Qy 501 SGSDPDNLIV---HSRPRQKHLHANGVPACTSK---LTKSLPASSTSDFRTHSCV 554
Db 1974 SPSASPSNLSLKSKQKAPLEDEIPS-----PLVKSRRVSFADPQQOQETADDIRRSPCI 2028
Qy 555 SEHSISVLNTPPEESKPSVARESTDQKMSVQSRPSREGCYS----GCSS 602
Db 2029 RTSSPRKPRNASSSQPKISEMSQE-----PRPVSRDCVYPALVGCSSA 2070

RESULT 9

US-10-369-493-22080
; Sequence 22080, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 22080
LENGTH: 1679
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22080

Query Match 7.4%; Score 234.5; DB 4; Length 1679;
Best Local Similarity 19.6%; Pred. No. 3.4e-06;
Matches 145; Conservative 124; Mismatches 230; Indels 241; Gaps 28;

QY 15 ENKNLSQYT-----SETKSPSSLYSQVLCSSVPLSKNVHGVGVCTGE--- 60
Db 1010 ENKDLSELVIRLEKDAADCOAELTKTKSSLYSAQDL-----LDKHERKWMEEKADYEREL 1064
QY 61 --NIEOS-----ISYLDQELTTFG-----FPSLYEESKSEAKRELNIYA 98
Db 1065 ISNIEOTESLRVNSVLIEKVDDTAANNQDKOHLKLVSLFNSURHERNSLETK----- 1117
QY 99 VLNCMNELLVQRNLLAQSIV-ETQNLKLGSDMDHLQSCYAKLKEQLETSRREMITG--- 154
Db 1118 LITCKRELAFVKQKNSLEKTINDLQRTQTLSEKEY--QCSAVIIDEFKDIKEVTQVMI 1175
QY 155 -----LQERDQLOCKNRSLLHLLKNEKDEKQVQKQNTIASRATQV---NHDVKRERE 204
Db 1176 LKENNAILOKSLKNVTEKREIYKQLNDROEETSRQLDQIQKEQVINSKNILVYSE 1235
QY 205 YNKLKERLHQLVWKKKNIAMDVLNVVGADKRGKSWRTDKTARNEDEMYKILLNDYE 264
Db 1236 MEQCKORYODLSQQQND-----AOKDIEKLTNE-----ISDLK 1269
QY 265 YRQKQILMENAELKVLQMKKEMISLSPQKKPRERAEDGTGTVAISDIEDSDGELSR 324
Db 1270 GKLSAENANADLENKFNRLKKAHEKLDASKQ----- 1303
QY 325 DSVWGLSCDTRBOLTNSIRKQWRILKSHVEKLDNOA-----SKVHSEGLNEED 373
Db 1304 -----QAALTNELNE-----LKAIKDKLEQDLHFENAKVIDLDTKLKAHELQSED 1348
QY 374 VISRODHEOET-----EKLELEIERCKE-MIKAAQQOL 404
Db 1349 V--SRDHEKDTVTLMEIEESLKKELQIFKTANSSSDAFEKLKVNMEKEDRIIDERTKE 1406
QY 405 LQQLATTCCDDTTSLRDCYLLLEEKRLKEW-----TL--FKEQKKNPERRERSPT 455
Db 1407 FEKKQETLNKSTSS---EAEYSKDLETLKKEWLKEYEDETLLRIKAEENLKKRIRLPS 1463
QY 456 EAIR--LGLERKAFEEERASWVKQ----- 479
Db 1464 EERIQIISKREELKEEFKLLKENAGSLTFLDNKSGEDAEELWNSPSKNSRPSA 1523
QY 480 ---FLNMTVPDRQNS-ENVKLFSAPSGS-----SDPDNLIVHSRPRQKLKLVANGVPAC 530
Db 1524 VAGFINQKRLKPOEQLKVNKNDVSNFNSQSMWTKENNIVDSAGNKAIPTFSGFKPFF 1583
QY 531 TSKLTKSLPASPTSDFROTHSCVSEHSSISVLNITPEE-----SKPSEVARESTDQ 582
Db 1584 SSN-TSSLQS--FQNPFTASQSNINTNAPRLTNIQPEVAVKAAINFNSVTDLTNNSTDG 1640
QY 583 KWSVQRSPSSRECGYSGCS 602
Db 1641 AKITEIGTSKRPISGCTSS 1660

RESULT 10
US-10-767-701-38884
; Sequence 38884, Application US/10767701
; Publication No. US20040172694A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 38884
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C77766_1.pep
US-10-767-701-38884

Query Match 7.4%; Score 233; DB 4; Length 206;
Best Local Similarity 31.5%; Pred. No. 2.8e-07;
Matches 57; Conservative 43; Mismatches 71; Indels 10; Gaps 4;

QY 54 GVFCGTGENTEQSISYLDQELTTFGPPSPSYEESKSEAKRELNIYAVLNCMNELLVLQKN 113
Db 29 GAFANAGNLEHCARYLNQLVTFGPPA-----SLDLFATDPVSTARTCNCIYALQQQRD 84
QY 114 LIAQESVETQNLKLGSDMDHLQSCYAKLKEQLETSRREMITGQERDQROCKNRSLLHQL 173
Db 85 IFRESTDNDQRMQSDISRLEAKIERMDAQLAAKDRRELATLTRTEAKNTAALKSQIDKL 144
QY 174 KNEKDEKQVQKQNTIASR--ATQYNHDVKREREYNKLKERLHQLVWKKDKN--IAMDVL 229
Db 145 QERDEFFQK--VIGNQVQRTQIHEMKKKEKEYIKQSKLNQVLMKKKSSRSRGWEIM 202
QY 230 N 230
Db 203 N 203

RESULT 11
US-10-437-963-180170
; Sequence 180170, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 180170
; LENGTH: 987
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_77560C.1.pep
US-10-437-963-180170

Query Match 7.4%; Score 233; DB 4; Length 987;
Best Local Similarity 22.4%; Pred. No. 2.1e-06;
Matches 151; Conservative 108; Mismatches 220; Indels 194; Gaps 35;


```
Db 462 DESVCSESVFNTLDLSIENWPNATKLNQ-----ENIESE 499
QY 66 ISYLDQELTFGPPSL---YEESKSEAKRELNI VAVLNCMBELLVLQKNLLAQ-ESVE 121
Db 500 LNSLRAD-----YDNLVLDYEQURTEKEEMELKL-KEKNDLDEFEALERKTKKQDESTE 553
QY 122 -----TONKLGSMDHDLQSCYAKL-----KEQLET 147
Db 554 DPKOMQKTLFPAETVALDAKRESAFURSENLELKEKMKELATTYKQWENDIQYQSLEA 613
QY 148 SRREMIGLOER-----DROLQCKNRSIHLHLLKNEKDEVOK 182
Db 614 KKKQWVDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLELGGKITDLOKELNKEVENE 673
QY 183 LQN--IIASRATQYNDHVKREREYNKLERLHOLVNMKKO-----NIAMDVLNY 231
Db 674 LREEVILLSLSEKSLPSEVERLRKEIQDKSEELH-IITSEKDKLFSEVVKHESRVQGLLEE 732
QY 232 VGRADKRGSRWTDKTEARNEDEMYKILLNDYEROKQILMENAELK-----279
Db 733 IGTKDDLATTSQYKSTQDFONFKTLHMDFEQYKMWLEENRNMQEI VNLKSBAQKF 792
QY 280 -----VLQOMKEMISLLSPKKPRERA-----303
Db 793 DSSLGALKTSLVKTQLOEKTREVOERLNEMLQLEQLENRDSPLQTVREKTLITEKL 852
QY 304 -----EDGTGTVAISDIE--DQSGELSRDSVWGLSCDTV-----REQLTNSIRKQWRIL 350
Db 853 QOTLEEVKTLTQEKDQLQLESQIERDQLKSDIHDVTVMNIDTQELRNALES---L 908
QY 351 KSHVEKLDNQASKVHSE-----GLNEE-----DV 374
Db 909 KOHOETINTLKSISEVSRNLHMEENTGETKDEFOOKVMGIDKKQDLEAKNTQTLTADV 968
QY 375 ISRQHEQOE-----TEKLEIEIERCKEM-----397
Db 969 KDNELIEQORKFLSLQEKNELOOMLESVIAEKEQLKTDLKENIEMTIENQELRLLGDE 1028
QY 398 IKAQOQLLOQ-----QIATTCDDTTSLRDYLLLEEKERLKEEWTLFKE-----442
Db 1029 LKKQBEIVAQEKNAHAIKKEGELSRTCD-----RLAEVEELKEKSQQLOEQKQOL 1078
QY 443 -----QKK-----NFERERSFTEAARLGLERKAFEEBRA 473
Db 1079 LNVQEMSEMQRKINSIENLKNELKNKELTLHEMETERLELAQKLMENYEEVKSIKERR 1138
QY 474 SHVQOFLNWTNFDH-----QNSENVKLFSAFSGSDPDNLI VHSRPRQKL 520
Db 1139 VLKELQKSFETERDHLRGVIREIATGLQTKELKIAH-----IHLKEHQTI 1186
QY 521 HSVANGVPACTSKLTKSLPASPTSDFRQTHSCVSEHSSISVLN-----ITPEESKPS 574
Db 1187 DELRRSVSEKTAQII-----NTQDLESHTKQLQE--EIPVLHBEQELLPNVKKVSE 1235
```

RESULT 14

```
US-10-408-765A-254
; Sequence 254, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408, 765A
; CURRENT FILING DATE: 2003-04-04
```

```
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: PASTSEQ for Windows Version 4.0
; SEQ ID NO 254
; LENGTH: 3225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-254
```

```
Query Match 7.0%; Score 222.5; DB 4; Length 3225;
Best Local Similarity 18.9%; Pred. No. 4.8e-05;
Matches 139; Conservative 147; Mismatches 217; Indels 233; Gaps 32;
```

```
QY 15 ENKULSQYTSFKMS-----PSSLYSQVLCSSVP-LSKNVHGVPFGFTGENIEQ 64
Db 1191 ENENIGDQURLOQIQVRESIDGKLPT--DQESCSSTPGLEEPFKATEQHHTQPVLES 1248
QY 65 SI-----SYLDQELTFFGPPSLVE---ESKSEAKRELNI VAVLNCMBELLVLQKNLLA 116
Db 1249 NLCPDWPSPHSDASALOGGTSVAQIKAQKETEAEKVELELKVSTTSSELTKKSEVFQL 1308
QY 117 QBSVETQNLKLS--DMHLSQCYAK-LKEQLETSRRMIGLQERDRLOQCKNRSIHLHLL 173
Db 1309 QEOINKQGLEIESLKTVSHEAEVHAESLQCKLESSQLIAGL-EHLRELQPK-----1359
QY 174 KNEKDEVOKLNIASRATQYNDHVKREREYNKLERLHOLVNMKKDNKNIAMDVLNYVG 233
Db 1360 ---LDELOKL-----ISKKEEDSVYLSGQ-----SEKAAALTKI---1391
QY 234 RADGKRGSRWTDKTEARNEDEMYKIL-----LNDYEROKQILMENAELK-----278
Db 1392 -----QTEIEQEDLIKALHTOLEMNAQEHDERIKOLQVELCEMKQKPEIGE 1439
QY 279 --KVLQOMKEMISLLSPQK-----KKPRERAEDGTGV-----AISDIED--DSGELS 323
Db 1440 ESRAQOIQORKLQAALISRKALKENKSLQELSARGTIERLTSLADVESQVSAQNK 1499
QY 324 RDSVWG-----LSCDTVR-----EQL 339
Db 1500 KDTVLGRLLALQERDKLITEMDRSLLENQSLSSCESLKLALLEGITDEKELVKEIESL 1559
QY 340 TNS-----IRKOWRIKLSHVKEKLDNOASKVHS--EGLNEB-----DVISR 377
Db 1560 KSKSTAESTEWKHEKLEQKEYEILLQSYENVSNAEARIQHVVEAVRQEKQELYGKLRST 1619
QY 378 QDHEQTEK---LELEIERCKEMIK---AQOQLLOQ-----LATTTC 413
Db 1620 EANKKETEQLOEAQEQEMEEKMKRPAKSKQKILELEEBEENDRLRAEVHPAGDTAKEC 1679
QY 414 DDDTTSLRDYLLLEEKERLKEEWTLFKQCKNKNFERERSFTEAARLGLERKAFEEBRA 473
Db 1680 ME--TLLSSNASMKBELEERVKMEYETLSKKFQSLMSSEKDSLSEEVQDLKHQIEGNVSKQA 1737
QY 474 SW-VKQOFLNWTNFDHNSENVKLFSAFSGSDPDNLI VHSRPRQKLHLSVANGVPACTS 532
Db 1738 NLEATEKHQDNQTNVTEGTQSIIP-----GETEQDSLSMSTR-----PTCSE 1779
QY 533 KLTKSLPASPTSDFRQTHSCVSEHSSIS-----VLNITPEESKPSFVARES 579
Db 1780 SVPSAKSANPAVSKDF-----SSHDEINNYLQIQDQLKRIAGLEEBEKQKNKFSQTL 1832
QY 580 TDQKWSVQSRPSSREG 595
Db 1833 ENKNTLLSQISTKDG 1848
```

RESULT 15

```
US-10-363-493-22285
; Sequence 22285, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
```

APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 22285
LENGTH: 1875
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22285

Query Match 7.0%; Score 220.5; DB 4; Length 1875;
Best Local Similarity 20.2%; Pred. No. 3.2e-05;
Matches 134; Conservative 119; Mismatches 238; Indels 171; Gaps 27;
QY 16 NKNLSQYTSKMSPLSYQ-QVLCSS---VPLSKNVHGVGVFCT----- 58
Db 1107 NSRIEDLSQNKL-----LYDQIQIYTAADKEVNNSTNGPLNNIILITRRERDILDTKVT 1162
QY 59 -----GENIEQISYLDQELTTFGPSPLYEESK-----EAKRELNIVA---VLNCM 103
Db 1163 VAERDAKMLRQKISLMDVEL-----QDARTKLDNSRVEKENHSSIIQOHHDDIMEKL 1213
QY 104 NELLVLQKXLLAQSVEQNLK---LGSMDHLQSCYAKLKEQLETSRREMIGLQERDR 160
Db 1214 NQNLRLRESNITLRNELENNNNKKELQSELQDKLQNVAPIESLTALKYSM---QEKEQ 1270
QY 161 QLOCKNRSLOHLLKDEKQVOKLQNIASRATQYNHDKRERYNKLKERLHQLVMNKK 220
Db 1271 EKLAKKEVHRKWSQDILEHEQLSS-----DYEKLESEIENLKEELE---NKE 1319
QY 221 DKNI-AMDVNTYVGRADGRGSRWTDKTEARNEDEMYKILLNDYBYRQKQIILMENAEK 279
Db 1320 RQGAEEKFNLRQAQER--LKTSLQSDSLTEQVNSLRDAKNVLENSLSEANARIEE 1377
QY 280 VLQMKKEMISLLSPQKKPRERAEDGTGTVAISDIEDSGELSRDSVWGLSCDTVR-EQ 338
Db 1378 -LQNAKV-----AQGNQLEAIRKLOEDAERAKSRELOAKLEESTTSYES 1420
QY 339 LTNSIRKOWRIKSHVEKLDNQASKVHSEGLNEEDVISRQDHEQTEKLELEIERCKEMI 398
Db 1421 TINGLNEEITTLKEIEKQRIQQOQATSANEQNDLS---NIVESMKKSFEEDKIK-FI 1476
QY 399 KAOQQLLOQQLATTCDDDTTSLRLDCYLLERKEERKEEWTL-FKEQKKNFERER----- 451
Db 1477 KEKTQEVNEK-----ILEAQERLNQPSNINMBEIKKKWESEHEQEVSQ 1519
QY 452 -----RSFTEAAILGLERKAFEEERASWVKQPLAMTNFDHNSENVKLF 498
Db 1520 KIRAEAEALKKRIPLTEKINKIIRKKEELEK-----EFEEKVEERIKSME 1567
QY 499 AFGSSD---PDNLIVHSRPROKHLHSHVANGVPACTSKLTKSLPASPS----- 545
Db 1568 Q-SGEIDVLRQLKQAKQKQKQKLENEYN-----KKLOEELKDVPHSSHISDDERDKL 1620
QY 546 -----DPROTHSCVSEHS-----SISVLNITPEESKPFSEVARESTQKQSVQSRP 590
Db 1621 RAEIESRLREEFNNELQAIKKKFDGKQQAAMKTTLLERKLAKMESQLSETKQSAESPP 1680
QY 591 SS 592
Db 1681 KS 1682

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 12, 2006, 19:16:24 ; Search time 9 Seconds
(without alignments)
863.952 Million cell updates/sec

Title: US-10-644-084-2

Perfect score: 3165

Sequence: 1 MGDWMTVPVLCTENKLS.....CYSGSSAFSAHGDRDDL 615

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 64916 seqs, 12643201 residues

Total number of hits satisfying chosen parameters: 64916

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New.*

- 1: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 2: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 3: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 4: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 5: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 6: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 7: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 8: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	343	10.8	382	US-10-953-349-6160	Sequence 6160, Ap
2	258	8.2	284	US-10-953-349-6161	Sequence 6161, Ap
3	213	6.7	232	US-10-953-349-6162	Sequence 6162, Ap
4	206.5	6.5	3113	US-10-505-928-325	Sequence 325, Ap
5	205	6.5	972	US-10-480-962-4	Sequence 4, Appli
6	201	6.4	621	US-11-293-697-4409	Sequence 4409, Ap
7	188	5.9	498	US-10-953-349-20047	Sequence 20047, A
8	188	5.9	508	US-10-953-349-20046	Sequence 20046, A
9	188	5.9	544	US-10-953-349-20045	Sequence 20045, A
10	187	5.9	824	US-11-293-697-4573	Sequence 4573, Ap
11	185	5.8	2871	US-10-505-928-100	Sequence 100, App
12	181	5.7	561	US-10-953-349-20041	Sequence 20041, A
13	178.5	5.6	486	US-10-953-349-24039	Sequence 24039, A
14	177	5.6	1234	US-11-327-900-6	Sequence 6, Appli
15	176	5.6	950	US-11-293-697-4492	Sequence 4492, Ap
16	173.5	5.5	541	US-10-953-349-17059	Sequence 17059, A
17	173	5.5	512	US-11-293-697-4080	Sequence 4080, Ap
18	171	5.4	450	US-10-953-349-24040	Sequence 24040, A
19	169.5	5.4	440	US-10-953-349-24041	Sequence 24041, A
20	169	5.3	519	US-11-293-697-4747	Sequence 4747, Ap
21	168	5.3	944	US-11-293-697-3086	Sequence 3086, Ap
22	165.5	5.2	515	US-10-953-349-17060	Sequence 17060, A
23	164.5	5.2	531	US-10-953-349-20042	Sequence 20042, A
24	164	5.2	513	US-10-953-349-11130	Sequence 11130, A
25	164	5.2	602	US-10-953-349-11129	Sequence 11129, A

ALIGNMENTS

RESULT 1

US-10-953-349-6160
; Sequence 6160, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 6160
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-953-349-6160

Query Match 10.8%; Score 343; DB 6; Length 382;
Best Local Similarity 25.8%; Pred. No. 1.1e-10;

Matches 95; Conservative 92; Mismatches 139; Indels 42; Gaps 10;

QY	56	FTGENIEOSISYLDQELTTGFPSPSYEESKSKAKRELINIVAVLNCNMELLVLQKRNLL	115
DB	28	FADVNLNLCIKYLNQSLVTSQFSA----SLDLFATDPVSIARTCNVYALIQQRQDVE	83
QY	116	AQESVETONLKLGSMDHLQSCYAKLKSQLETSRREMIGLQERDRLOQCKNLSLHOLLKN	175
DB	84	FRESANDQRQLLSDMARLEAKVERLETQLQKEREGLSVTTEAKNTAALTKQEKJQK	143
QY	176	ENDEVQKLQNIAS--ATQYNHVDYKREYNKLERLHLQVLM--NKDKDNIAVDVNYV	232
DB	144	ERDEFQRM--VIANQVQTKQLHETKKEKEYIKLQERLQVLMKKETRSQMEIMNLL	201
QY	233	GRADKGRSWRTDKTEARNEDEMYKILLNDVEYRQKILMENAELKVLQOMKEMISLL	292
DB	202	QKEGRQRTWSGKKTDS----DFYKKIIVDAEAKNQLMAENTDLRALLRSTQGMRSFL	257
QY	293	SPQKKKPRERADG-----TGTVAISDIEDSGELSRDSVWGLSCDTVREOLTNSJR	344
DB	258	N-----ASGGUTNOSLVANGRHGADPSQPLG--GKTDVFDLPFRMARGQLEDSJR	306
QY	345	KQWRIKSHVEKLDNQAKVHSEGLNEEDVISRQHEQETKLELEIIBRCKEMIKAQOOL	404
DB	307	SKWYSIKERMGLVDAQKEVS-----ITSEASERELE-LEAQLVEARSIIQOESI	356
QY	405	LOQLATT	412

Db 984 SLNQKNNLIQSEFANYIDEREKSTSELSDOYKQEKIL-LQRCETGTNAVYEDLSQY 1042
QY 408 QLATTCCDDTTSLRCYLLLEERLKEEWTLFKEQKKNFERRRSFTFAAIBLGLERKA 467
Db 1043 KAAQKNNKLECLNCTSLCENRQNELE-----QLKEAFKHEQEF-----LTKLA 1089
QY 468 FEERASWVK-----QQFLNMTNFQHNSENKVLFSAFSGSSDPDNLIVHSRPROKKLH 521
Db 1090 FAERNQNLMLLETVQALRSEMTDNQNSK-----SEAGLQKEIMTLKEEQNKMQ 1142
QY 522 SVANGVPACTSKLTKSLPASPTSDPROTHSCVSEHSSISVLNITPEESKPSSEVARESTD 581
Db 1143 KEVNDLLQNEQLMKVM-----KTKHEC-----QNLSESPIN--- 1175
QY 582 QKWSQSRPSSREGC 596
Db 1176 ---SVKERESERNQC 1187

RESULT 5

US-10-480-962-4
; Sequence 4, Application US/10480962
; Publication No. US20060115813A1
; GENERAL INFORMATION:
; APPLICANT: TANG, Y. Tom; BAUGHN, Mariah R.;
; APPLICANT: LEE, Ernestine A.; RAMKUNAR, Jayalaxmi;
; APPLICANT: YUE, Henry; GRIFFIN, Jennifer A.;
; APPLICANT: CHAMLA, Narinder K.; TRAN, Bao
; APPLICANT: GANDHI, Ameena R.; HAPALIA, April J.A.;
; APPLICANT: SWARNAKER, Anita; GURURAJAN, Rajagopal;
; APPLICANT: POLICKY, Jennifer L.; YAO, Monique G.;
; APPLICANT: WARREN, Bridget A.; GRIETZEN, Kimberly J.;
; APPLICANT: ELLIOTT, Vicki S.; LEE, Soo Yeun;
; APPLICANT: SANJANWALA, Bharati; HONCHELL, Cynthia D.;
; APPLICANT: FORSTHE, Ian J.; GORVAD, Ann E.;
; APPLICANT: RICHARDSON, Thomas W.; LEE, Sally;
; APPLICANT: BARROSO, Ines

; TITLE OF INVENTION: INTRACELLULAR SIGNALING MOLECULES

; FILE REFERENCE: PF-1002 USN

; CURRENT APPLICATION NUMBER: US/10/480,962

; CURRENT FILING DATE: 2003-12-04

; PRIOR APPLICATION NUMBER: PCT/US02/17955

; PRIOR FILING DATE: 2002-06-06

; PRIOR APPLICATION NUMBER: US 60/297,010

; PRIOR FILING DATE: 2001-06-08

; PRIOR APPLICATION NUMBER: US 60/298,706

; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/299,998

; PRIOR FILING DATE: 2001-06-20

; PRIOR APPLICATION NUMBER: US 60/300,377

; PRIOR FILING DATE: 2001-06-21

; PRIOR APPLICATION NUMBER: US 60/301,871

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/303,349

; PRIOR FILING DATE: 2001-07-05

; PRIOR APPLICATION NUMBER: US 60/303,403

; PRIOR FILING DATE: 2001-07-06

; PRIOR APPLICATION NUMBER: US 60/351,927

; PRIOR FILING DATE: 2002-01-25

; NUMBER OF SEQ ID NOS: 40

; SOFTWARE: PERL Program

; SEQ ID NO 4

; LENGTH: 972

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID No: 3418524CD1

US-10-480-962-4

Query Match

6.5%; Score 205; DB 6; Length 972;

Best Local Similarity 19.8%; Pred. No. 0.0021;
Matches 126; Conservative 111; Mismatches 245; Indels 154; Gaps 24;

QY 63 EOSISYLDQELTTFGPPSLYEESKKEAKRELNIIVAVLNCMNEELLVLRKNLLAQSSET 122

Db 176 ESKLSSMNSIKTFMSPEL-----KKERALKDEASKITIKWEQYRVVQEENQHWMTQIA 231

QY 123 --ONLKGSDMHL-----QSCYAKLKE-----QLETSSREMIGLQERDRL 162

Db 232 LODELRQIDLQNLFOQDSSRTGEPVCAELTEENFORLHABHERQAKELFLRLKLEEM 291

QY 163 OCKNRSLLHLLKNEKDEVOKLQNIITAS-----RATQYNHD-----VK 199

Db 292 ELRIETQKOTLNARDESIKKLLEMLQSKLSAKATEEDHERTRRLAEAEHVHLLSLDE 351

QY 200 RKEREYNKLERLHQLVNMNKDKNIAMDVLNVVGRADKRGSGWRTDKTKEARNDEMYK-- 257

Db 352 QKEKENSMLREEMHRRFENAPDSAKTQALQTVIEMKDSKISSMERGLRDLDEEIIQMLKSN 411

QY 258 --ILLNDYEROKQILMENAE---LKKVLQMKKEMISLLSPQKKPRERAEDGTGTVAI 312

Db 412 GALSTEEREEMKQMEVYRSHSKFMKNKVEQLKEE---LSSKEAQWELKKKAAGLQA- 466

QY 313 SDIEDDSGELSDSVWGLSCDTRVROLTNSIRKQWILKSHVEKL-----DNQASKVH 365

Db 467 -EIGQVKQELSRKOTELLALQTKLETN-----QFSDSKOHIEVLKESLTAKEQRAAILQ 521

QY 366 SE-----GLNEEDVI---SRQDHEQETEK-----LELEIE 392

Db 522 TEVDALRLRLLEKETMLNKKTKIQIDMAEEKTQAGEIHDLDKMDLVKERNVNVLOKKIE 581

QY 393 RCKEMIKAAQQLLQO-----QLATCDDTTSLLRDCYLLLEEKERLKEEWTLFKEQK 444

Db 582 NLQEQLRDKKEQKQSSILKERVKSLOADTTNTDTALTTLLEA--LAEKERTIER---LKEQR 636

QY 445 KNFERE-----RRSFTEAAIRLGLERKAFEEERASW--VKQOFLNMTNPDHONSIV 494

Db 637 DRDEREKQEEIDNYKDKLKEKVSLLQGDLSKEASLLDLKEHASSLASSGLKKXDSRL 696

QY 495 KLFSAFSGSSDPDNLIVHSRPRQKKLHVSANGVPACTSKLTKSLPASPTSDPROTHSCV 554

Db 697 KTLLEALEQKKEECLKMS--QLKKAHEAA-----LEARASPEMSDRIQ----- 738

QY 555 SEHSSISVLNITPEESKPS-----EVARESTDQK 583

Db 739 --HLEREITRYKDESSKQAQAEVDRLLEILKEVENEK 772

RESULT 6

US-11-293-697-4409
; Sequence 4409, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4409
; LENGTH: 621
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-293-697-4409

Query Match 6.4%; Score 201; DB 7; Length 621;

Best Local Similarity 19.2%; Pred. No. 0.0019;

Matches 97; Conservative 115; Mismatches 182; Indels 110; Gaps 20;

QY 12 LCTENKNSLQYTSYSETYKMSFSSLYSQVLCSSVPLSKNVHGVFGVCTGENIEQSISYLDQ 71


```

; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 24039
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Glycine max
US-10-953-349-24039

```

[illegible]

```

RESULT 14
US-11/321-900-6
; Sequence 6, Application US/11327900
; Publication NO. US20060099174A1
; GENERAL INFORMATION:
; APPLICANT: PHARMA PACIFIC
; APPLICANT: Meritex Jean Francois
; APPLICANT: Dron, Michel
; APPLICANT: Tovey, Michael Gerard
; TITLE OF INVENTION: INTERFERON-ALPHA INDUCED GENE
; FILE REFERENCE: 46658/250044
; CURRENT APPLICATION NUMBER: US/11/327,900
; CURRENT FILING DATE: 2006-01-09
; PRIOR APPLICATION NUMBER: US/10/203,311
; PRIOR FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: GB 0002979.3
; PRIOR FILING DATE: 2000-02-09
; PRIOR APPLICATION NUMBER: GB 0002980.1
; PRIOR FILING DATE: 2000-02-09
; PRIOR APPLICATION NUMBER: GB 0002982.7
; PRIOR FILING DATE: 2000-02-09

```

```

, PRIOR APPLICATION NUMBER: GB 0002981.9
, PRIOR FILING DATE: 2000-02-09
, NUMBER OF SEQ ID NOS: 8
, SOFTWARE: PatentIn Ver. 2.1
, SEQ ID NO 6
, LENGTH: 1234
, TYPE: PRT
, ORGANISM: Homo sapiens
US-11-327-900-6

```

Query Match	5.6%	Score 177;	DB 7;	Length 1234;
Best Local Similarity	20.8%	Pred. No. 0.066;		
Matches 133;	Conservative 103;	Mismatches 231;	Indels 172;	Gaps 28;
Qy	108	VLQKNLLAQESVETQNLKLGSDMDHLQSCYAKLKEQLETSRREM----	IGLOERDRQLQ	163
Db	585	IIQENERLQOEILEKN-----RIEQNDKISELIERNQRYVEOSNLMMEKRNNSLQ		636
Qy	164	CKNRSLHQLLKNKEDEVQKQLNIASRATQYNDHVKRKEREYNKLERLHQLVMNKDKDN	223	
Db	637	TATENTQARVLHAEQKAKVTELAATAQVSH-----LQLKMTAQH-----KLETE	683	
Qy	224	IAMDVLNYYGRADGRGSM----RTDTEARNEDEMYKILLNDYERQOKILMENAEKLV	280	
Db	684	LQMWLTESLKETDLLRGQLTKVQAKLSELQETSEQAQSKPKSEKQNRKQLELKVTSLEEE	743	
Qy	281	LQMKKEMISL----LSPQKKK----PRERADGTGTVAISDIE----DSGEISLRDSVWGLSC	332	
Db	744	LTDRLRYEKESLEKNLSERKKKSQAQERSQABEEIDEIRKSYQEBELDKLRQLLKKT--RVST	801	
Qy	333	D-TVREQLT--NSIRKQWRILKSHVKEKLDNQAKVHSEGLNE--EDVISQD--HEQETEK	386	
Db	802	DQAAEQSLSVQAELOTWEE-----AKCEHLLASAKDEHLQOYQEVCAQORDAYQOKLVQ	855	
Qy	387	LELEIERC-----KEMIKAQOQLLQOQLATTCDDDTT-----S	419	
Db	856	LQ--EKCLALQAOITALTKQNEHQIKLEKKNKSONSGVEAASDPSEKVKYKIMNQVQS	912	
Qy	420	LIRD-----CYLLEEKERLKEEWTLFKEQKKNFRERRRSFT	455	
Db	913	LRREFELEESYNGRTILGTIMNTIKMVTQLLNNQOQEKEEESSEBEEKAEERPRRPSQ	972	
Qy	456	E--AATRLGLERKAFEEH-----ASVWKKQOFLNMTVFDHQ--NSENVKL	496	
Db	973	EQSASASSQOQAPLNRPESGPMVPEQVVEAVPLPQALTTSDGHRKKDGEAEAL	1032	
Qy	497	FSAFSSGSDPDNLIVHSRRPRQKKLHSHVAN-----GVPACTSKLTKSLPASPTS-	545	
Db	1033	SEIKGSLPPELSCIPS-----HRVLGPPTSIPPELGPVSMDSCEESLAASPMNAK	1085	
Qy	546	-DFRQTHSCVSE-----HSSISVLNIT--PEESKPSVEARESTDQKMSVQSRP----	590	
Db	1086	PDNPSKVCVREAVDPGLQESSTRLSLTSDFPEGDPLALGPSPGPEPPQLKDDVTS	1145	
Qy	591	-----SSREGCVSGCASSAFPSAH-----GORDD	613	
Db	1146	STGPHKELSTAGSTVGAAGALRPSHHSQSRSSLSLSDGEED	1184	

```

RESULT 15
US-11-293-697-4492
; Sequence 4492, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ IDS NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1

```

```
/ SEQ ID NO 4492
/ LENGTH: 950
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-11-293-697-4492

Query Match      5.6%; Score 176; DB 7; Length 950;
Best Local Similarity 19.4%; Pred. No. 0.054;
Matches 124; Conservative 105; Mismatches 231; Indels 178; Gaps 26;

QY 60 ENIEQISYLDQELTFGPPSPSYESKSKAEKRELNIVAVLNCMNELLV-----LQRKN 113
Db 65 ERAEQILRATEEF-----KQLEEAQLKKISEAGKDLLYKQLSGRLQLVN 110
QY 114 LLAQESVETQNLKLGSDMDHLQSCYAKLQELTSSREMIGLQERDROLQCKNRSL---- 169
Db 111 KLRQALD-----LQLQEKQKQEIAGKQEIKDQLAIDSLDSKD 151
QY 170 --HQLLKNEK-----DEVQKLQNIISRAQYQNHVKKRERYNKLKERL--HQLVMN 218
Db 152 PKHSHMKAQKSGEQQLDIMNKYQQLSRLDEILSRIAKETEEIKDLEEQLTEGQIAAN 211
QY 219 ----KUD-KNIAMDVINYVG-----RADKKGSRWTK-----TEARNEDENMYKILL 260
Db 212 EALKKDLEGVISGLQELGTIRKQATQACNECKLRDEKETILLQRLTEVEQERDQLEIVA 271
QY 261 NDVEYRQOI-----LMENAEKKVLQOMKEM-----ISLLSPQKKKPRERAE 304
Db 272 MDAENMRKELAEALQEHVNASLQQTQGDLSAYAELEARNLNRDAEANQLKEBLE 331
QY 305 DGTGTVAISDIEDDS--GELSRDSVWGLSGCDTVREQLTNSIRKQW----- 347
Db 332 KVT---RLTQLSGLQAELEKE-----RQALKNALGKAQFSEKQENSELHAK 378
QY 348 -----RILKSHVEKLDNQASKVHSEGLNEEDVISQDHEQETEKL----ELEIERCKE 396
Db 379 LKHLQDDNNLLKQQLKDFQNLHNLHVVDGLVRPEEVAARVDELRRKCLKGTGEMNIHSPSD 438
QY 397 MIKAOQQLLQQLATTCCDDTTSLARDCYLLEE---KERLKEWTLFKEQKKNERERR 452
Db 439 VLGKSLADLQKQFSEI-----LARKWERDEAOVRERKKQEEEMALQOEKLATQOEFR 491
QY 453 SPTAAI--RLGLERKAFEEERASWVKQFLNMTNFDHQNSENVKLFSAFSGSSD----- 505
Db 492 QACERALEARNWFDKQHE-----ARIQOMENEIHVLOENLAKSMEEIQGLTDLQOE 543
QY 506 ----PDNLIVSR--PRQKHLHVN-----GVPACTSKLTKSLPASPSSTSDPROT---H 551
Db 544 ADEEKERILAQURELEKKKKLEDAKSQEQVGLDKELKKLKKAVATSKLATAEITIAKD 603
QY 552 SCVSEHSSISVLNITPEES-----KPSEVAREST 580
Db 604 QLSLHGTVKINQERAELEQAEERFSRKAQAARDLT 641
```

Search completed: June 12, 2006, 19:19:29
Job time : 11 secs

GenCore version 5.1.9
Copyright (C) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 13, 2006, 23:11:04 ; Search time 891 Seconds
(without alignments)
7218.742 Million cell updates/sec

Title: US-10-644-084-2

Perfect score: 3165
Sequence: 1 MGDWMTVPVLCITENKNLS.....CYSGSSAFRSAGHRDDLP 615

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlp
-Q=/abss/ABSSWEB_spool/US10644084/runat_12062006_150410_1340/app_query.fasta_1
-DB=N Geneseq -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss03p
-USER=US10644084 @CGN_1_1147 @runat_12062006_150410_1340 -NCPU=6 -ICPU=3
-NO_MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=3 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_8.*

1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*
15: Geneseqn2006s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length DB	ID	Description
1	3165	100.0	2692	12 ADO26490	Ado26490 Murine af
2	2937	92.8	3195	12 ADO26492	Ado26492 Rat afadi
3	2725.5	86.1	5476	10 ACD19444	Acd19444 CDNA enco

SUMMARIES

	4	2720	85.9	2272	10	ADB63343
5	1982.5	62.6	2716	4	AAH14625	Adh63343 Human CDN
6	1192	37.7	826	4	AAH06926	Aah14625 Human CDN
7	856	27.0	1503	4	AAH16996	Aah16996 Human CDN
8	854	27.0	668	13	ADO78551	Adg78551 Novel can
9	715	22.6	563	4	AAH08911	Aah08911 Human CDN
10	671.5	21.2	498	2	AAV86670	Avv86670 EST clone
11	650	20.5	403	5	AAF66702	Aaf66702 Novel hum
12	635.5	20.1	428	3	AAO3006	Aao3006 Human sec
13	380.5	12.0	1691	13	ADX36698	Adx36698 plant ful
14	359.5	11.4	1556	4	ABE67285	Aeb67285 Rice geno
15	353	11.2	1544	3	ACA43689	Aca43689 Zea maye
16	343	10.8	1149	3	ACA46459	Aca46459 Arabidops
17	334.5	10.6	1062	13	ADR64454	Adr64454 Cotton CD
18	328	10.4	513	12	ACH72399	Ach72399 Human gen
19	326.5	10.3	1496	3	AAO33585	Aao33585 Arabidops
20	317	10.0	1485	3	ACA49413	Aca49413 Arabidops
21	310	9.8	1044	3	ACA49412	Aca49412 Arabidops
22	308.5	9.7	833	4	AAH03373	Aah03373 Human CDN
23	290	9.2	1199	13	ADX27288	Adx27288 Plant ful
24	285	9.0	178	12	ACH86155	Ach86155 Human gen
25	234.5	7.4	5040	13	ADT47329	Adt47329 Bacterial
26	225.5	7.1	7509	14	ADM44256	Adm44256 Human cen
27	224	7.1	2138	14	AEA51203	Aea51203 Chicken R
28	224	7.1	2943	3	AAH70229	Aah70229 Plasmodiu
29	223.5	7.1	10300	6	ABK70274	Abk70274 Human lun
30	223.5	7.1	10300	13	ADZ44554	Adz44554 Human bre
31	223.5	7.1	10300	14	ADZ26156	Adz26156 Human gen
32	223.5	7.1	10300	14	ABD18268	Aed18268 Fibrotic
33	222	7.0	6619	5	ABA20636	Abas20636 Human ner
34	222	7.0	6619	5	ABA20638	Abas20638 Human ner
35	222	7.0	6619	5	ABA20637	Abas20637 Human ner
36	220.5	7.0	5628	13	ADT47534	Adt47534 Bacterial
37	220.5	7.0	6284	5	ABX71393	Abx71393 Human cel
38	220.5	7.0	6284	13	ADX97475	Adx97475 Pancreat
39	220.5	7.0	6284	15	AEH84913	Aee84913 IMPDH mar
40	220.5	7.0	6386	4	AAI57837	Aai57837 Human poi
41	219.5	6.9	2602	4	ABL29009	Abi29009 Drosophil
42	218	6.9	4747	13	ADQ87627	Adq87627 Human tum
43	218	6.9	7962	3	AAZ58980	Aaz58980 Human cyt
44	218	6.9	8347	6	ABK13313	Abk13313 DNA enco
45	218	6.9	8347	12	ADQ85937	Adq85937 Human tum

ALIGNMENTS

RESULT 1
ADO26490
ID ADO26490 standard; DNA; 2692 BP.
XX
XX ADO26490;
XX
XX 29-JUL-2004 (first entry)
XX
XX Murine afadin dilution domain binding protein (ADIP) DNA SeqID 1.
XX murine; mouse; gene; ds; afadin dilution domain binding protein; ADIP;
XX afadin; actinin; binding inhibitor; cardiac; heat disease;
XX myocardial infarction; myocarditis.
XX Mus musculus.
XX
XX Key Location/Qualifiers
XX CDS 80..1927
XX FT /*tag=
XX FT /product= "ADIP protein"
XX
XX JP2004135658-A.
XX
XX 13-MAY-2004.
XX
XX 14-AUG-2003; 2003JJP-00293554.
XX

PR 27-SEP-2002; 2002JP-00284263.
 XX (EISA) EISAI CO LTD.
 PA
 XX
 XX WPI; 2004-404616/38.
 DR P-PSDB; ADO26491.
 XX
 New polynucleotide encoding an afadin dilution domain binding protein
 PT having avidity with afadin or actinin, useful for diagnosing heart
 PT diseases e.g. myocardial infarction.
 XX
 PS Claim 1; SEQ ID NO 1; 37pp; Japanese.
 XX
 CC This invention relates to a novel isolated nucleic acid encoding an
 CC afadin dilution domain binding protein (ADIP) that exhibits an avidity
 CC with afadin/actinin. Specifically, it refers to screening assays to
 CC identify compounds that modulate ADIP avidity and provides suitable
 CC agonists, antagonists and antibodies thereof. The present invention
 CC provides methods to identify afadin and actinin binding inhibitors
 CC therapeutically as cardiants to diagnose and/ or treat heart disease such
 CC as myocardial infarction or myocarditis. This polynucleotide sequence is
 CC the murine ADIP DNA sequence of the invention.
 XX
 SQ Sequence 2692 BP; 814 A; 525 C; 691 G; 662 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,96e-243 Length: 2692
 Scores: 3165.00 Matches: 615
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 12 Gaps: 0
 US-10-644-084-2 (1-615) x ADO26490 (1-2692)
 QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCyThrGluAsnLysAsnLeuSer 20
 DB 80 ATGGAGATTTGGATGACTGTGACAGATCCAGTCTGTGTACAGAAAACAAAAATCTCTCT 139
 QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
 DB 140 CAATATACCTCAGAAACAAAGATGTCCTCGTCCAGTTTGTATCTCCAGCAAGTTCTGTGC 199
 QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
 DB 200 TCTTCAGTACCTTTATCCAAAACGTCGATGGTGTTCCTCGGTGCTTCTGACAGGAGAG 259
 QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
 DB 260 AACATTGAACAAAGTATTTCTCTATCTTGATCAGGAGCTGACCACCTTCGGGTTCCTTCC 319
 QY 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
 DB 320 TTGTATGAAGAAATCCAAAGTAAAGAGCAAGAGAGAAATTAATATATAGTCGCTGTTCTG 379
 QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
 DB 380 AACTGTATGAACGAGCTGCTGCTGCTTCAGCGGAGAACCTGCTGGCCACGAGAGCGTG 439
 QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnInSerCysTyrAlaLys 140
 DB 440 GAGACACAGAACTTGAAGCTGGCAGTGACATGGACACCACTTCGACAGCTGCTACGCCAAA 499
 QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
 DB 500 CTTAAGGAGCAGTTGGAACGTCACGCGGAGAGATGATCGGGCTTCAAGAGAGAGACAGG 559
 QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
 DB 560 CAGCTGCAGTGCAGAACACAGGAGTTTGATCAGCTCTCTGAAGATGAGAAAGATGAGGTA 619
 QY 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200

DB 620 CAAAAAATTACAAAAATATCATAGCCAGCGGGCTACTCAGTATAATCATGATGTGAAGAGG 679
 QY 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
 DB 680 AAGGAGCGTGANTATATAATAGCTAAGAGCGCCCTGCATCAGCTCGTTTATGAACAAGAG 739
 QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
 DB 740 GATAAAAACATAGCCATGGATGTTTTAAATTTATGTGGTCCGAGCTGATGGCAACAGGCG 799
 QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
 DB 800 TCATGGAGGACTGACAAAACAGAGCCAGGAATGAAGATGAGATGTACAAAATTTCTGTGTG 859
 QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
 DB 860 AATGATTATGATGATACCCGACAGACAGATCTCTGATGGAGAACGGGAGCTGAAGAGGTC 919
 QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
 DB 920 CTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTCTCTCTCAGAGAAGAGGCCAGG 979
 QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
 DB 980 GAAAGAGCAGAGGACGGCAGGACCTGTTGCTATCTCCGATATAGAGATGATGACTCTGGG 1039
 QY 321 GluLeuSerArgAspSerValTyrGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
 DB 1040 GAACCTGAGCAGACAGACGGTGTGGGGCTTTCTCTGTACACTGTGAGAGAGCAGCTGACA 1099
 QY 341 AsnSerIleArgLysGlnTyrArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
 DB 1100 AACAGCATCAGGAACACAGTGGAGAAATTTGAAAGTTCATGTAGAAAAAATCGATAACCAA 1159
 QY 361 AlaSerLysValHisSerGluGlyLeuAsnGluAspValIleSerArgGlnAspHis 380
 DB 1160 GCTTCGAAGGTACACTCAGAGGGCTTTAATAGAGAGGACGTCATCTCACGACAAGACCAT 1219
 QY 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400
 DB 1220 GAGCAAGAGACTGAGAAACTGGAGCTGGAGATTGAGCGGTGTAAGAGATGATCAAGGCT 1279
 QY 401 GlnGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
 DB 1280 CAGCAGCAGCTCTTTACAGCAGCAGCTGCCACCACTGATGATGATGATGATGATGATGATGAT 1339
 QY 421 LeuArgAspCysTyrLeuLeuGluLysGluArgLeuLysGluTyrThrLeuPhe 440
 DB 1340 TTGGCAGACTGTACTTGTCTGGAAGAAAGAGAACGCTTTAAAGAGAGGTGGACCTTTT 1399
 QY 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
 DB 1400 AAAGAGCAAAAAAAGAAATTTGAGAGAGAAAGGCGAAGCTTTACAGAAGCTGCCATTGCA 1459
 QY 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPhe 480
 DB 1460 TTGGGGTTGGAGAAAAGGGCTTTTGAAGAAAGAGCGACGCTGGGTAAAGCAGCAGTTT 1519
 QY 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
 DB 1520 TTTAAACATGACGAACTTTGACCCACCAACTCAGAAAATGTGAAACTTTTTCAGTGCCTTC 1579
 QY 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
 DB 1580 TCAGGAAGTTCTGATCCAGACAATCTTATAGTCCACTCAGCGCCACGCAAGAGAGCTA 1639
 QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
 DB 1640 CACAGTGTGGCTAATGGGGTGCACAGTTCACATCAAAACTGACTAAATCTCTTCTCTGCC 1699
 QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
 DB 1700 TCACCTTCTACTTCAGACTTTTCGCCAGACACATTCATGTGTGTGTAACACAGATTCATC 1759

QY 561 SerValLeuAsnIleThrProGluSerLysProSerGluValAlaArgGluSerThr 580
 DB 1760 AGTGCTGAATATAACTCTGAAGAAAGTAAACCAAGTGAGGTTCCAAAGAGAAACGACG 1819
 QY 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGlyCys 600
 DB 1820 GATCAGAAGTGGAGCTGCAGTCGAGCCCGCCAGCTCGCGGGAGGGTGCTACAGCGGATGC 1879
 QY 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
 DB 1880 TCCTCGGCTTCAGGAGCGCTCACGGGACCGAGATGACTTACCT 1924

RESULT 2

ADO26492

ID ADO26492 standard; DNA; 3195 BP.

XX AC ADO26492;

XX 29-JUL-2004 (first entry)

XX DE Rat afadin dilution domain binding protein (ADIP) DNA SeqID 3.

XX rat; gene; ds; afadin dilution domain binding protein; ADIP; afadin;

XX KW actinin; binding inhibitor; cardiant; heat disease;

XX KW myocardial infarction; myocarditis.

XX OS Rattus norvegicus.

XX FH Key Location/Qualifiers

XX FT 79..1920

XX FT /*tag= a

XX FT /product= "ADIP protein"

XX PN JP2004135658-A.

XX PD 13-MAY-2004.

XX PF 14-AUG-2003; 2003JP-00293554.

XX PR 27-SEP-2002; 2002JP-00284263.

XX PA (EISA) EISAI CO LTD.

XX DR WPI: 2004-404616/38.

XX DR P-PSDB; ADO26493.

XX PT New polynucleotide encoding an afadin dilution domain binding protein
 PT having avidity with afadin or actinin, useful for diagnosing heart
 PT diseases e.g. myocardial infarction.
 XX Claim 1; SEQ ID NO 3; 37pp; Japanese.
 XX CC This invention relates to a novel isolated nucleic acid encoding an
 CC afadin dilution domain binding protein (ADIP) that exhibits an avidity
 CC with afadin/actinin. Specifically, it refers to screening assays to
 CC identify compounds that modulate ADIP avidity and provides suitable
 CC agonists, antagonists and antibodies thereof. The present invention
 CC provides methods to identify afadin and actinin binding inhibitors
 CC therapeutically as cardiants to diagnose and/ or treat heat disease such
 CC as myocardial infarction or myocarditis. This polynucleotide sequence is
 CC the rat ADIP DNA sequence of the invention.

XX SQ Sequence 3195 BP; 929 A; 648 C; 794 G; 823 T; 0 U; 1 Other;
 Alignment Scores:
 Pred. No.: 4,45e-225 Length: 3195
 Score: 2337.00 Matches: 567
 Percent Similarity: 95.1% Conservative: 18
 Best Local Similarity: 92.2% Mismatches: 28
 Query Match: 92.8% Indels: 2
 DB: 12 Gaps: 1

US-10-644-084-2 (1-615) x ADO26492 (1-3195)

QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
 DB 79 ATGGGAGATTGGATGACTGTTACAGATCCAGTCTTGTTACAGAAAAACAAAATCTCTCT 138
 QY 21 GlnTyThrSerGluThrLysMetSerProSerSerLeuTySerGlnGlnValLeuCys 40
 DB 139 CAATATACCTCAGAAACAAAGATGTCCTCCGTCAAGTTTATATCTCGCAGCAAGTACTGTGC 198
 QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
 DB 199 TCTGCAACACCTTTATCCAAAGATGTGCATGGTGTTCCTCAGTCCTCTTCACAGAGAG 258
 QY 61 AsnIleGluGlnSerIleSerTyrlleuAspGlnGluLeuThrPheGlyPheProSer 80
 DB 259 AACATCGAACACAGATATTTCGTATCTTGATCAGGAACCTGACTACCTTCGCTTCCTTCC 318
 QY 81 LeuTyGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
 DB 319 TTGTATGAAGAAATCCAAAAGTAAGGAGCGCAAGCAGAGTAAAGTATAGTTGCTCTCTG 378
 QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
 DB 379 AACTGCATGATGAGCTGCTTGTGCTTCAGCGAAGAACCTCTCTGCCCCAGGNAACGGTG 438
 QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrlAlaLys 140
 DB 439 GAGACACAGNATCTGAAGCTGGGCAGTGACATGGACCACCTGCAGAGCTGCTACGCCAAA 498
 QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
 DB 499 CTTAAGGAACAGTTGGAGGCCCTCCAGCGCAGAGATGATCAGCCTTCAGAGAGAGACAGA 558
 QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180
 DB 559 CAGCTACAGTGCNAAAACAGGAATTTGCATCAGCTCTCTGAANAACGAGNAAGAGGTG 618
 QY 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrlAsnHisAspValLysArg 200
 DB 619 CAAAAATTTACAAAATATCATAGCCAGTCGGGCTACTCAGTATAATCATCATGTAAGAGA 678
 QY 201 LysGluArgGluTyrlAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
 DB 679 AAGGAGCGGGAGTACAATAAACTGAAGGAGCGCTGTCATCAGCTTGTATTGAACAAGAAG 738
 QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrlValGlyArgAlaAspGlyLysArgGly 240
 DB 739 GATAAAAATATAGCCATGACCGTCTTTAAATTACGTGGCCGAGTGGATGGAAAGCGAGGC 798
 QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrlLysIleLeuLeu 260
 DB 799 TCCTGGAGGACTGATAAAACAGAGCCAGGAATGAAGATGAATGTACAAAATTTCTGCTG 858
 QY 261 AsnAspTyrlGluTyrlArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysVal 280
 DB 859 AATGATTATGATGATACCGCAGAGCAGATCTCTGCGAGAAATGCGGAGCTGAAGAGGTC 918
 QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
 DB 919 CTCCAGCAGATGAGAGAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 978
 QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
 DB 979 GAAAGAGCAGAGGACAGCACAGGCACTGTTGTGTCATCTCCGATGTAGAAGACGACGCTGG 1038
 QY 321 GluLeuSerArgAspSerValTrpClyLeuSerCysAspThrValArgGluGlnLeuThr 340
 DB 1039 GAGCTGACAGCAGATGGTGTGGAGCCCTTTCTCTGTGACCTGTCTCAGGAGCAGCTTACA 1098
 QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
 DB 1099 AACAGCATCAGGAGCAGTGGAGAAATTTCTGAANAAGCCATGTGGAAAAAATCTGTAACCA 1158

```
QY 361 AlaserLysValHisSerGluGlyLeuAenGluGluAspValIleSerArgGlnAepHis 380
Db 1159 GTTTCARAGGTACATCTCAGAGGCTTTCATGAAGAGGAGCTCATCTCAGCACAAACCAT 1218
QY 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400
Db 1219 GAGCAAGAGACTGAGAACTGGAGCTGGAGATTGAGCGGTGTAAGAGATGATCAAGGCT 1278
QY 401 GlnGlnGlnLeuGlnGlnGlnLeuAlaThrThrCysAepAspAspThrThrSerLeu 420
Db 1279 CAGCAGCAGCTCTTACAGCAAAGCTGGCCACTGCGTGTGATGAGCAGCACACCTCACTG 1338
QY 421 LeuArgAspCysTyLysLeuGluLysGluArgLeuLysGluGluGluTrpThrLeuPhe 440
Db 1339 TTGCGAGACTGTACTTCTGCTGAAGAAAGAACGCCCTTAAGAGAGAGTGGTCCCTTTT 1398
QY 441 LysGluGlnLysLysAspPheGluArgGluArgArgSerPheThrGluAlaAlaIleArg 460
Db 1399 AAGAGCAAAAAAGAAATTTTTCAGAGAGAAAGACCAAGCTTTACAGAAAGCTGCTATTGCG 1458
QY 461 LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnGlnPhe 480
Db 1459 TTGGGGTGGAGAGAAAGCGTTTTCAGGAAGAGCGCAGCGCTGGGTGAAGCAGCAGTTT 1518
QY 481 LeuAenMetThrAsnPheAspHisGlnAenSerGluAenValLysLeuPheSerAlaPhe 500
Db 1519 TTAAACATGACGACCTTTGATCACCAGAACTCAGAAATGTGAATCTTTCAGTGCCTTT 1578
QY 501 SerGlySerSerAspProAspAenLeuIleValHisSerArgProArgGlnLysLysLeu 520
Db 1579 TCAGAAAGTTCTGATCCAGACAATCTTATAGTCCACCCACGCCACGGCAAAAGAGCCA 1638
QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db 1639 CACATGTGCTGCTAAATGGGGTGCCAGCTTGCACATCCAACTGCTTCTTCCGACC 1698
QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
Db 1699 TCACCT-----TCAGACTTCTGCCCGCTCGCTCATGTGTCTGAGCAGCAGTCCCGTC 1752
QY 561 SerValLeuAenIleThrProGluSerLysProSerGluValAlaArgGluSerThr 580
Db 1753 AGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1812
QY 581 AspGlnLysTrpSerValClnSerArgProSerSerArgGluGlyCysTyLysSerGlyCys 600
Db 1813 GACCAAGAGTGGAGCGTGTGCTCCAGACCCAGCTCCCGGAGGGTGTCTACGTTGGATGC 1872
QY 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615
Db 1873 TCCTCGGCTACACAAGCTCCACCCTGGAGACGAGATGACTTACCA 1917
RESULT 3
ID ACD19444
AC ACD19444 standard; cDNA; 5476 BP.
XX AC ACD19444;
XX XX
DT 25-AUG-2003 (first entry)
DE cDNA encoding novel human protein #124.
KW Human; NOV; gene therapy; endocrine related disease; diabetes;
KW Metabolism-related disease; obesity; central nervous system disorder;
KW Alzheimer's disease; Parkinson's disease; epilepsy; multiple sclerosis;
KW schizophrenia; depression; autoimmune disorder; inflammatory disorder;
KW psoriasis; allergy; lupus erythematosus; asthma; cancer;
KW inflammatory bowel disease; rheumatoid arthritis; osteoarthritis;
KW colon cancer; lung cancer; liver cancer; breast cancer; ovarian cancer;
KW prostate cancer; brain cancer; melanoma; liver disease; liver cirrhosis;
KW lung disease; emphysema; obstructive pulmonary disease; haemophilia;
KW stroke; infection; drug; ss.
```

```
XX OS Homo sapiens.
XX PN WO2003023002-A2.
XX PD 20-MAR-2003.
XX PF 09-SEP-2002; 2002WO-US028539.
XX 07-SEP-2001; 2001US-0318120P.
XX 07-SEP-2001; 2001US-0318130P.
XX 10-SEP-2001; 2001US-0318430P.
XX 17-SEP-2001; 2001US-0322636P.
XX 17-SEP-2001; 2001US-0322781P.
XX 17-SEP-2001; 2001US-0322816P.
XX 17-SEP-2001; 2001US-0322817P.
XX 19-SEP-2001; 2001US-0323519P.
XX 20-SEP-2001; 2001US-0323631P.
XX 20-SEP-2001; 2001US-0323636P.
XX 25-SEP-2001; 2001US-0324969P.
XX 25-SEP-2001; 2001US-0325091P.
XX 26-SEP-2001; 2001US-0324980P.
XX 17-APR-2002; 2002US-0373212P.
XX 06-SEP-2002; 2002US-00236177.
XX (CURA-) CURAGEN CORP.
XX SPytek KA, Patturajan M, Gorman L, Li L, Anderson DW, Zhong M;
XX Shinkach VL, Vernet CAM, Ellerman K, Berghs C, Rothenberg ME, Guo X;
XX Rieger DM, Taupier RJ, Shenoy SG, Liu X, Padigar M, Alebrook JP;
XX Lepley DK, Edinger SR, Burgess CE;
XX WPI: 2003-313242/30.
XX P-FSDB; ABO14751.
XX New cytoplasmic, nuclear membrane bound or secreted polypeptides (NOVX)
XX and polynucleotides, useful in gene therapy, e.g. for treating or
XX preventing obesity, multiple sclerosis, allergy, cancers, hemophilia,
XX stroke or infections.
XX Claim 20; Page 351-353; 586pp; English.
XX The invention describes a new isolated polypeptide (NOVX). The NOVX
XX polypeptide, nucleic acid and antibody are useful as therapeutics,
XX particularly in the manufacture of a medicament for treating a syndrome
XX associated with a human disease, which includes a pathology associated
XX with NOVX polypeptide. The DNA encoding the protein is useful in gene
XX therapy for treating the disease or condition. In particular, the NOVX
XX polypeptide or polynucleotide is useful for treating endocrine/
XX metabolism-related diseases (e.g. obesity or diabetes), central nervous
XX system disorders (e.g. Alzheimer's disease, Parkinson's disease,
XX epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune
XX and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,
XX asthma, inflammatory bowel disease, rheumatoid arthritis or
XX osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,
XX prostate or brain cancers, or melanoma), liver diseases (e.g. liver
XX cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),
XX haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).
XX These are also useful in developing powerful assay system for functional
XX analysis of various human disorders, as well as in diagnostic
XX applications, and for monitoring the effects of drugs during clinical
XX trials. This sequence encodes a novel human NOV protein
XX Sequence 5476 BP; 1806 A; 887 C; 1063 G; 1720 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 7,73e-208 Length: 5476
Score: 2725.50 Matches: 527
Percent Similarity: 93.1% Conservative: 34
Best Local Similarity: 87.7% Mismatches: 39
Query Match: 86.1% Indels: 1
DB: 10 Gaps: 1
```

US-10-644-084-2 (1-615) x ACD19444 (1-5476)

15 GluAenLysAenLeuSerGlnTyrThrSerGluThrLysMetSerProSerSerLeuTyr 34
127 GAAAGCAAAACCTATCTCTCAATATATACCTCAGAAACAAAGATGCTCCATCAAGTTTATAC 186
35 SerGlnGlnValLeuCySerSerValProLeuSerLysAenValHisGlyValPheGly 54
187 TCACAGCAAGTGCTATGTTCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTCAGT 246
55 ValPheCyThrGlyGluAenIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThr 74
247 GCCTTCTGCACAGAAGATAATATCAACACAGAGTATCTCATATCTTGTGATCAGGAATTGACT 306
75 ThrPheGlyPheProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeu 94
307 ACTTTTGGTTTTCTCTTCAATATATGAAGAATCCAAAGGTAAGAGACAAAGACAGAGATTA 366
95 AenIleValAlaValLeuAenCysMetAenGluLeuLeuValLeuGlnArgLysAenLeu 114
367 AATATATGATGCTGACTAATTTGATGATGAGCTGCTTGTGCTTCAGCGGAAGAACCTT 426
115 LeuAlaGlnGluSerValGluThrGlnAenLeuLysLeuGlySerAspMetAspHisLeu 134
427 CTAGCTCAGGAAATGTGCAGACACAGAAATTTGAAGCTGGGAAGTGATATGACCATCTA 486
135 GlnSerCyThrAlaLysLeuLysGluGlnLeuGluThrSerArgArgGluMetIleGly 154
487 CAGAGCTGCTACTCAAACTTTAAGGAACACTCGGAACCTCCAGAGGAGAAATGATGGG 546
155 LeuGlnGluArgAspArgGlnLeuGlnCysLysAenArgSerLeuHisGlnLeuLeuLys 174
547 CTTTCAGGAAAGAGACAGACAGTTCACATGTGAACACAGAAATTTGCATCAGCTACTAAG 606
175 AsnGluLysAspGluValGlnLysLeuGlnAenIleAlaSerArgAlaThrGlnTyr 194
607 AATCAGAAAGATGAGGTGCAAAAATTAACAAATATCATTTGCAAGTCGAGCTACTCAGTAT 666
195 AsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGln 214
667 AATCATGATATGAAGAGAAAGAGCGTGAAATATAATAAATCTGAAGGAACGCTACATCAA 726
215 LeuValMetAenLysLysAspLysAenIleAlaMetAspValLeuAsnTyrValGlyArg 234
727 CTTGTTATGAACAGAAAGATAAGAAATAGCTATGGACATTTTGAATTTATGTCGGGAGA 786
235 AlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAspGlu 254
787 GCTGATGGAANAAGAGGCTCTCGGAGACTCGTAAACTGAAGCCAGGAATGAAGATCAA 846
255 MetTyrLysIleLeuAenAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAen 274
847 ATGTATAAAAATCTCTTGAATGATATGAATATCGTCAGAAACAAATCTCTAATGGAAAAT 906
275 AlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerPro 294
907 GCAGAACTTAAAGAGTTCTTCAACAAATGAANAAGGNAATGATTTCTCTCTTTCTCC 966
295 GlnLysLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIleSerAsp 314
967 CAANAAGAAGAAACCTAGAGAAAGATGATGATAGTACAGGAACTGTTTCTCTCTCTCC 1023
315 IleGluAspAspSerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThr 334
1024 GTTGAAGAGATGCCGGGAACTAAGCAGAGAGATGATGTGGGACCTTTCTCTGTGAAACT 1083
335 ValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisVal 354
1084 GTGAGAGAGAGCTTACAAACAGCATCAGAAAAACAGTGGAGAAATTTGAAAGTCATGTA 1143
355 GluLysLeuAspAenGlnAlaSerLysValHisSerGluGlyLeuAenGluAspVal 374

```
FT FT /*tag= a
XX /product= "Clone TESTI20071130 protein"
PN EPI308459-A2.
XX
XX
XX 07-MAY-2003.
XX
XX 28-MAR-2002; 2002EP-00007401.
XX
XX 05-NOV-2001; 2001JP-00379298.
XX 25-JAN-2002; 2002US-00350978.
XX
XX (HELI-) HELIX RES INST.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
XX WPI; 2003-450961/43.
XX P-PSDB; ADB65313.
XX
XX New polynucleotides and polypeptides, useful for developing a diagnostic
XX marker or medicines for regulation of their expression and activity, or
XX as targets of gene therapy.
XX
XX Claim 1; Page; 222pp; English.
XX
XX The invention discloses a polynucleotide comprising a sequence selected
XX from 1970 fully defined nucleotide sequences which encode novel
XX polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
XX or its partial peptide, an antibody binding to the polypeptide or peptide
XX of the polynucleotide, immunologically assaying the polypeptide or
XX peptide of the polynucleotide by contacting the polypeptide or peptide
XX with the antibody of the encoded protein, and observing the binding
XX between the two, a transformant carrying the polynucleotide in an
XX expressible manner and an antisense polynucleotide. The oligonucleotide
XX is useful as a primer for synthesising the polynucleotide, or as a probe
XX for detecting the polynucleotide. The polynucleotides and encoded
XX proteins are useful as pharmaceutical agents and many disease-related
XX genes may be included in them, for developing a diagnostic marker or
XX medicines for regulation of their expression and activity, or as targets
XX of gene therapy. The genes are involved in tissue and/or cell
XX regeneration. Membrane proteins, signal transduction-related proteins,
XX transcription-related proteins, disease-related proteins and genes
XX encoding them can be used as indicators for diseases (e.g. osteoporosis,
XX neurological diseases, cancer, tumours). The cDNA may be used to regulate
XX the activity or expression of the encoded protein to treat diseases. The
XX sequence presented is a cDNA of the invention. Note: Some of the sequence
XX data for this patent is not represented in the printed specification, but
XX is based on sequence information supplied by the European Patent Office.
XX
XX Sequence 2272 BP; 783 A; 398 C; 502 G; 589 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 6,87e-208 Length: 2272
XX Score: 2720.00 Matches: 536
XX Percent Similarity: 86.8% Conservative: 34
XX Best Local Similarity: 81.6% Mismatches: 40
XX Query Match: 85.9% Indels: 47
XX DB: 10 Gaps: 3
XX
XX US-10-644-084-2 (1-615) x ADB63343 (1-2272)
XX
XX QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
XX DB 224 ATGGAGAGATGGATGACTGTTACAGATCCA-----GAAAGCAAAACTATCTCT 271
XX
XX QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
XX DB 272 CAATATACCTCAGAAACCAAGATGCTCCATCAAGTTTATCTACTACACCAAGTGTATGT 331
XX
XX QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
XX
Db 332 TCCTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTTCAGTGCCTTCTGCACAGAAGAT 391
QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
Db 392 AATATTGAACAGAGTATCTCATATCTTGTATCAGAAATTGACTACTTTTGGTTTCCCTTCA 451
QY 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db 452 TTATATGAGAGATCCAAAGGTTAAGAGACAAGAGAGAGATTAAATATATAGTAGCTACTA 511
QY 101 AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
Db 512 AATTGTATGAATGAGCTGCTTGTGCTTCAGCGGAAGAACCTTCTAGCTCAGGAAATGTG 571
QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisIleuGlnSerCysTyrAlaLys 140
Db 572 GAGACACAGAATTTGAAGCGGGAAGTATATGACCACTCTACAGAGCTGTACTCAGAAA 631
QY 141 LeuLysGluGlnLeuGluThrSerArgGluMetIleGlyLeuGlnGluArgAspArg 160
Db 632 CTTAAGGNACAACTGGAAACCTCCAGGAGGAATGATTGGCTTCAGGAAAGAGACAGA 691
QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
Db 692 CAGTTACAATGTAAGAACAGGAATTTGCATCAGCTACTTAAAGAAATGAGAAAGATGAGGTG 751
QY 181 GlnLysLeuGlnAsnIleIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Db 752 CAAAAATTTACAAAAATATCTTGCAGTGCAGCTACTCAGTATATATCATGATGTAAGAGA 811
QY 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db 812 AAAGAGGTGATATATATAAATCTGAAGGAGCTCTACATCAACTTGTATGACAAAGAAA 871
QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db 872 GATAAGAAAAATAGCTATGGACATTTGAAATTTATGTCGGGAGAGCTGATGGAAGAGAGGC 931
QY 241 SerTrpArgThrAspLysThrGluAlaArgGlnAspGluMetTyrLysIleLeuLeu 260
Db 932 TCCTGGAGGAGCTGTGTAACACTGGAACCCAGGAAATGAAGATGAATGTATATAAATTTCTGTG 991
QY 261 AsnAspTyrGluTyrArgGlnLysGlnLeuMetGluAsnAlaGluLeuLysLysVal 280
Db 992 AATGATTATGAATATCGTCAGAAACAATCCTAATGGAAATGCAGAACTTAAGAAGGTT 1051
QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysArgProArg 300
Db 1052 CTTCAACAAATGAAAAAGGAATGATTTCTCTTCTTCTCCCAAGAAAGAAACCTAGA 1111
QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSergly 320
Db 1112 GAAAGAGTAGATGATAGTACAGGAACCTGTT---ATTTCCGATGTTGAAGAGATGCCGGG 1168
QY 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
Db 1169 GAACCTAAGCAGAGAGATATGTGGACCTTTCTGCTGTGAACCTGTGAGAGAGCAGCTTACA 1228
QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
Db 1229 AACAGCATCAGAAAAACAGTGGAGAATTTTGAAGAGTCATGTAGAAAAACCTTGATATACCAA 1288
QY 361 AlaSerLysValHisSerGluGluLeuAsnGluGluAspValIleSerArgGlnAspHis 380
Db 1289 GTTTCAAGAGGTACACCTGGGAAGGTTTTAATGATGAAGATGTAATCTCAGCAAGACCAT 1348
QY 381 GluGlnGluThrGluLysLeuGluGluIleGluArgCysLysGluMetIleLysAla 400
Db 1349 GAACAAGAACTGAAAAACTCGAGTTAGAAATTCAGCAGCTGTAAGAAATGATTAAAAACT 1408
QY 401 GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
```

```
Db 1409 CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCATATGATGATGATACCATCTTCACTA 1468
Qy 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTyrThrLeuPhe 440
Db 1469 TTACGAGAGCTGTTATTTGTTGGAAGAAAGAAAGCACTCTCAAGAGAAGTGGTCCCTTTT 1528
Qy 441 LysGluGlnLysLysAsnPheGluArgGluGluArgSerPheThrGluAlaAlaLeuArg 460
Db 1529 AAGACGCGANAAAGAAATTTTGAGGGAGAGACGAAGCTTTTACAGAAGCCGCTATTGCG 1588
Qy 461 LeuGlyLeuGlu----- 464
Db 1589 CTGGGATTGGAGATTGGCTTTCTCTCGCCAAAGTGTTTTACACAAGAAATCTCTTTGGGT 1648
Qy 464 ----- 464
Db 1649 ATCCATCATTCATCCAGCTGAGTAGTGTTGTCAGAAGACAGTGCACCAAGACTAT 1708
Qy 465 -----ArgLysAlaPheGluGluArgGluArgAlaSerTrpValLysGlnG 479
Db 1709 ATTGCTGTTTAAAAAGAAAGGCATTTGAAGAAGAAAGAGCCAGTGGTTAAAGCAGCA 1768
Qy 479 nPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAl 499
Db 1769 GTTCTTAATATGACTACCTTTGACCACCAAGACTCAGAAATGTGAACTTTTCAGTGC 1828
Qy 499 aPheSerGlySerSerAspPheAspHisGlnAsnSerGluAsnValHisSerArgProArgGlnLys 519
Db 1829 CTTCTCAGGAAGTCTGATTGGGACAATCTTATAGTGCACCTCGAGGCGCGCAAGAA 1888
Qy 519 sLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuPr 539
Db 1889 GCCTCACAGTGTCTAAATGGGTCTCCAGTTGCGATGTCTAACTTACTAAATCTCTTCC 1948
Qy 539 oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe 559
Db 1949 TGCTTCACTTCCACTTCAGACTTTTGCCAGACACGCTTCTGCATATCTGAACATAGTTC 2008
Qy 559 rLysSerValLeuAsnLeuThrProGluGluSerLysProSerGluValAlaArgGluSe 579
Db 2009 AATCAATGTACTGAATATTAACCTGCTGAAGAAATTAACCAAAATCAGGTTGGAGGACG 2068
Qy 579 rThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGl 599
Db 2069 TACAATCAAAATGAGGTGGCGTCAAGACTGGATCAGAGACTGGATCAGGAGGTTGCTATAGTG 2128
Qy 599 yCysSerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615
Db 2129 ATGCTCTTGAGCTACACAAATCTCATGTAGAAAAGATGACTTACCT 2177
```

RESULT 5

```
AAH14625
ID AAH14625 standard; cDNA; 2716 BP.
AC AAH14625;
XX
DT 26-JUN-2001 (first entry)
DE Human cDNA sequence SEQ ID NO:12264.
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS Homo sapiens.
PN EP1074617-A2.
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 12264; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 2716 BP; 935 A; 405 C; 524 G; 852 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1e-148 Length: 2716
Score: 1982.50 Matches: 383
Percent Similarity: 92.6% Conservative: 27
Best Local Similarity: 86.5% Mismatches: 32
Query Match: 62.6% Indels: 1
DB: Gaps: 1

US-10-644-084-2 (1-615) x AAH14625 (1-2716)
Qy 173 LeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArgAlaThr 192
Db 3 CTAAAGATGAGAAAGATGAGGTGCAGAAATTTACAAATATCATTTGCAAGTGGAGTACT 62
Qy 193 GlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeu 212
Db 63 CAGTATATCATGATATGACAGAGAAAGAGCGTGAATATATAAATGAAGGACGCTCTA 122
Qy 213 HisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrVal 232
Db 123 CATCAACTTGTATGACAAGAAAGATAGCAAAATAGCTATGCGATTTTGAATTTATGTC 182
Qy 233 GlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgGlnGlu 252
Db 183 GGGAGAGCTGATGGAAAGAGGCTCTCGAGGAGCTGGTAAACCTGAAGCCAGGAATGAA 242
Qy 253 AspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMet 272
Db 243 GATGAATGTATAAAATTTCTCTTGAATGATATGATATCGTCAGAACAAATCTCTAATG 302
Qy 273 GluAsnAlaGluLeuLysLysValLeuGlnMetLysLysGluMetIleSerLeuLeu 292
```

```
Db 303 GAAATGCAAGCTTAAAGAGCTTCTCAACAAATGAAAGGAAATGATTTCTCTCTT 362
Qy 293 SerProGlnLysLysProArgGluArgAlaGluApoGlyThrGlyThrValAlaIle 312
Db 363 TCTCCCAAGAAAGAAACCTTAGAAGAAAGAGTAGTAGTACAGGAACGTT---ATT 419
Qy 313 SerAspIleGluAspSerGlyGluLeuSerArgAspSerValTptGlyLeuSerCys 332
Db 420 TCCGATGTTGAAGAGATGCCGGGAACTAACACAGAGAGATGTGGACCTTCTCTGT 479
Qy 333 AspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSer 352
Db 480 GAAACTGTGAGAGAGAGCTTACAAACAGCATCAGAAACAGCTGCAGAAATTTTGAAGA 539
Qy 353 HisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGluGlyLeuAsnGluGlu 372
Db 540 CATGTAGAAAGCTTGTATAACCAAGTTTCAAAAGGTACACCTGGAAAGGTTTAAATGATGA 599
Qy 373 AspValIleSerArgGlnAspHisGluGlnGluThrGluLysLeuGluGluIleGlu 392
Db 600 GATGTATCTCAGCAAGACCATGACAAAGAACTGAAAACTCGATTTAGAAATTCAG 659
Qy 393 ArgCysLysGluMetIleLysAlaGlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThr 412
Db 660 CAGTGTAAAGAAATGATTAAAACTCAGCAACAGCTTTTACAGCAGCAGCTGCCTACTGCA 719
Qy 413 CysAspAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLysGluArg 432
Db 720 TATGATGATGATACACCTTCACTATTTACGAGACTGTTATTTGTGGAAAGAAAGGACGT 779
Qy 433 LeuLysGluGluThrLeuPheLysGluGlnLysLysAsnPheGluArgGluArg 452
Db 780 CTCAAAGAGAAATGTGCTCCCTTTTAAAGAGCAGAAAGAAATTTTGAGAGGAGACAGA 839
Qy 453 SerPheThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArg 472
Db 840 AGCTTTACAGAGCCGCTATTCCGCTGGGATTTGGAGAGAAATGGCATTTTGAAGAAAGA 899
Qy 473 AlaSerTrpValLysGlnPheIleuAsnMetThrAsnPheAspHisGlnAsnSerGlu 492
Db 900 GCCAGTGTGTTAAAGCAGCAGTTTCTAAATATGACTTACCTTTTGACCCACAGAACTCAGAA 959
Qy 493 AsnValLysLeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHis 512
Db 960 AATGTGAAACTTTTCAGTGCCTTCTCAGGAAGTTCTGATTGGGACAAATCTATAGTGAC 1019
Qy 513 SerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCysThrSer 532
Db 1020 TCGAGGCGAGCCGCAAAAGAGCCTCACAGTGTGTCTAATGGGTCTCCAGTTTGCATGTCT 1079
Qy 533 LysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArgGlnThrHisSer 552
Db 1080 AAACCTTACTAAATCTCTCTGCTTCCCTTCCACTTCCACTTCCAGACTTTTGGCAGCACGTTCC 1139
Qy 553 CysValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGluSerLysPro 572
Db 1140 TGCATATCTGAACATAGTTTCATCATGTAATTAATTAATTAATTAATTAATTAATTAACCA 1199
Qy 573 SerGluValAlaArgGluSerThrAspGlnLysTrpSerValGlnSerArgProSerSer 592
Db 1200 AATCAGGTTGGAGGAGATGTACAAATCAAAATGGAGTGTGGCGTCAAGACCTGGATCA 1259
Qy 593 ArgGluGlyCysTyrSerGlyCysSerSerAlaPheArgSerAlaHisGlyAspArgAsp 612
Db 1260 CAGGAAGGTTGCTATAGTGGATGCTCCTTGGAGCTACAAATTTCTCATGTAGAAAAAGAT 1319
Qy 613 AspLeuPro 615
Db 1320 GACTTACCT 1328
```

RESULT 6
AAH06926

AAH06926 standard; cDNA; 826 BP.
AAH06926;
26-JUN-2001 (first entry)
Human cDNA clone (5'-primer) SEQ ID NO:3761.
Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
Homo sapiens.
EP1074617-A2.
07-FEB-2001.
28-JUL-2000; 2000EP-00116126.
29-JUL-1999; 99JP-00248036.
27-AUG-1999; 99JP-00300253.
11-JAN-2000; 2000JP-00118776.
02-MAY-2000; 2000JP-00183767.
09-JUN-2000; 2000JP-00241899.
(HELI-) HELIX RES INST.
Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
WPI; 2001-318749/34.
Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
Claim 1; SEQ ID NO 3761; 2537pp + Sequence Listing; English.
The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent polynucleotides, all of which are used in the exemplification of the present invention
Sequence 826 BP; 329 A; 120 C; 179 G; 195 T; 0 U; 3 Other;

Alignment Scores:
Pred. No.: 4,48-86 Length: 826
Score: 1192.00 Matches: 240
Percent Similarity: 92.7% Conservative: 14
Best Local Similarity: 87.6% Mismatches: 18
Query Match: 37.7% Indels: 4
Gaps: 1

US-10-644-084-2 (1-615) x AAH06926 (1-826)

```
QY 173 LeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleIleAlaSerArgAlaThr 192
Db 3 CTAAGAATGAGAAAGATGAGGTGCAGAAATATACAAATATATCATTCAGAGTACT 62

QY 193 GlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeu 212
Db 63 CAGTATAATCATGATGATGAAGAGAGAAAGCGTGATATAATAATAAAGTGAAGGACCTCTA 122

QY 213 HisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrVal 232
Db 123 CATCAACTTGTATGAACAAGAAGAATAAGAAATAGCTATGCACATTTTGAATTATGTC 182

QY 233 GlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrCysLysAlaArgAsnGlu 252
Db 183 GGGAGAGCTGATGGAAAAGAGCTCTCGAGAGACTGGTAAACCTGAAGCCAGGAATGAA 242

QY 253 AspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMet 272
Db 243 GATGAATGATATAAATCTCTTGAATGATTATGAATATCGTCAGAAACAATCTCTAATG 302

QY 273 GluAsnAlaGluLeuLysValLeuGlnGlnMetLysLysGluMetLysSerLeuLeu 292
Db 303 GAAATGCAAGACTTAAGAGGTTCTTCAACAATGAAAAGGAAATGATTTCTCTCTT 362

QY 293 SerProGlnLysLysLysProArgGluArgAlaGluAspGlyThrClyThrValAlaIle 312
Db 363 TCTCCCAAGAAGAAACCTAGAGAAAGATGATGATAGTACAGGAATGTT---AFT 419

QY 313 SerAspIleGluAspSerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCys 332
Db 420 TCCGATGTTGAAGAGATCCCGGGAACCTAAGCAGAGAGATATGTGGACCTTTCCTGT 479

QY 333 AspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSer 352
Db 480 GAAACTGTGAGAGAGCAGCTTACAAACAGCATCAGAAACAGTGGAGATTTTGAAGT 539

QY 353 HisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGluGlyLeuAsnGluGlu 372
Db 540 CATGTAGAAAAGCTGTGATAACCAAGTTTCAAGAGTACACCTCGAAGGTTTAAATGATGA 599

QY 373 AspValIleSerArgGlnAspHisGluGlnGlnThrGluLysLeuGluLeuGluIleGlu 392
Db 600 GATGTAATCTCAGCAGACAGCATGACACAGAAACTGAAAACCTCGAGTTAGAAATTCAG 659

QY 393 ArgCysLysGluMetIleLysAlaGlnGlnGlnLeuLeuGlnGlnLeuAlaThrThr 412
Db 660 CAGTGTAAAGAAATGATTAAACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCA 719

QY 413 CysAspAspAspThrThrSerLeuLeuArgAspCysTyrLeu-LeuGluGluLysGluArg 432
Db 720 TATGATGATGATCC-AC TTC-CTATTACGAGACTGTATTGTTGGAGAAAGGACCG 777

QY 432 gLeuLysGluGluTyrThrLeuPheLysGluGlnLysLys 445
Db 778 TNYTCAAGAAAATGGGCCCTTTTAAAGAACCCNAAAA 817

RESULT 7
AAH16996
ID AAH16996 standard; cDNA; 1503 BP.
XX
AC AAH16996;
XX

DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:16285.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
```

```
XX 28-JUL-2000; 2000EP-00116126.
PF
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
length cDNAs defined in the specification, and for the detection and/or
diagnosis of the abnormality of the proteins encoded by the full-length
cDNAs.
XX
PS Claim 8; SEQ ID NO 16285; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-
length cDNAs defined in the specification. Where a primer set comprises:
(a) an oligo-dr primer and an oligonucleotide complementary to the
complementary strand of a polynucleotide which comprises one of the 5602
nucleotide sequences defined in the specification, where the
oligonucleotide comprises at least 15 nucleotides; or (b) a combination
of an oligonucleotide comprising a sequence complementary to the
complementary strand of a polynucleotide which comprises a 5'-end
sequence and an oligonucleotide comprising a sequence complementary to a
polynucleotide which comprises a 3'-end sequence, where the
oligonucleotide comprises at least 15 nucleotides and the combination of
the 5'-end sequence/3'-end sequence is selected from those defined in the
specification. The primer sets can be used in antisense therapy and in
gene therapy. The primers are useful for synthesizing polynucleotides,
particularly full-length cDNAs. The primers are also useful for the
detection and/or diagnosis of the abnormality of the proteins encoded by
the full-length cDNAs. The primers allow obtaining of the full-length
cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
represent human amino acid sequences; and AAH13629 to AAH13632 represent
present invention
XX
SQ Sequence 1503 BP; 490 A; 269 C; 271 G; 473 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 7,72e-59 Length: 1503
Score: 856.00 Matches: 198
Percent Similarity: 44.0% Conservative: 9
Best Local Similarity: 42.1% Mismatches: 11
Query Match: 27.0% Indels: 253
Db: 4 Gaps: 1
```

US-10-644-084-2 (1-615) x AAH16996 (1-1503)

```
QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
Db 97 ATGGGAGATTGGATGACTGTACAGATCCAGGCTGTCTTCAGAAAGCAAACTACTCTCT 156

QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db 157 CAATATACCTCAGAAACAAGATGCTCCATCAAGTTTATATCTACAGCAAGTGTATGT 216

QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 217 TCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTTCAGTGCCTTCTGCACAGAAAT 276

QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGln----- 71
Db 277 AATATTGAACAGAGATATCTCATATCTTGATCA-GGTAACATGTTTTTTGTAAAAAACAGTA 335
```

QY 71 ----- 71
Db GCATAATCATGTTATATATAAGATGTATCAAAATTATGAGAAAAACATTTGTCAACA 395
QY 71 ----- 71
Db CAAGTTTGTGTTATGTAGTGCCCTGAGACTTAGAATTTTCAGAGAACCTTGGTGATTA 455
QY 71 ----- 71
Db TCTCATCTAGAGATAGCAAGTCTAACTTTTGTGTGGAGCACCTTGGCTCTCAATTGT 515
QY 71 ----- 71
Db CCCAGGGCTCTCCACTGAGAGTCCAGCCATCACTTTATGATGTCAGAAACCCAAAGTGC 575
QY 71 ----- 71
Db CTTGCTCCATGTCACAGATTTATAGCAGAGACTAAAGAGCCCGTCTCTTAACCTCTTT 635
QY 71 ----- 71
Db GTCCTTTGGGATTAATGACCAATTTCTTAACATATTTCAAGATAGTTTTCTATTTTCTAAG 695
QY 71 ----- 71
Db GAAAGATTCTAATCTTTTGAATTTTTTTTACTAAGTTGCCAGTAACCAACCTTTTAAA 755
QY 71 ----- 71
Db ATAATTCACATTTATTTTAAATTTTGACAAAAATTATATATATTTGTCAATGTTCAACAT 815
QY 71 ----- 71
Db AATGTTTCATAATACACATTTGTGAATGGCTCAGTGTGCATACCTCCCATCTATATCA 875
QY 71 ----- 71
Db TTTTGTGGTGAGAAATTTACTTGCTCAGCAATTTTCAGGAATATAGTAGTAGTATTAG 935
QY 71 ----- 71
Db CTGAGGTCACCGTGTGTACAATAGATCTATTAAACTTATTCTTCTATCTAAGTGAAG 995
QY 71 ----- 71
Db TTTGTATCTCTGACCACTATCTCTCAACTGTAAACCACTTTAAGAAAAACATTTTCTCC 1055
QY 72 -----GluLeuThrThrPheGlyPheProSerLeuTyrGluGluSerLysSerLysGlu 89
Db TACCAGGAATTGACTACTTTTGGTTTCTTCATTTATGAGATCCAAAGGTAAAGAG 1115
QY 90 AlaLysArgGluLeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuValLeu 109
Db ACNAAAGACAGAGTTAAATATAGTAGCTGTACTATAATTTGTATGAATGAGCTGTGCTT 1175
QY 110 GlnArgLysAsnLeuLeuAlaGlnGluSerValGluThrGlnAsnLeuLysGlySer 129
Db CAGCGGAAGAACCCTTCTAGCTCAGGAAAATGTGGAGACACAGAAATTTGAAGCTGGGAAGT 1235
QY 130 AspMetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuThrSerArg 149
Db GATATGGACCATCTACAGAGCTGTCTCAAAACTTTAAGGAACAACCTGGNAACCTCCAGG 1295
QY 150 ArgGluMetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeu 169
Db AGGGAATGATTTGGGCTTCAGGAAAGACAGACAGAGTTACAATGTAAAGACAGGAATTTG 1355
QY 170 HisGlnLeuLeuLysAsnGluLysAspGluValGln-LysLeuGlnAsnIleIleAlaSer 189
Db CATCAGCTACTAAGAANTGAGAAAGATGAGGTGCAAAAAAATTTACAAAATATATCATTTGCAAG 1415

QY 189 rArgAlaThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLys 209
Db TCGAGCTACTCAGTATATATCATGATATGAAGAAAGACGCTGAATATATAAATCTGAA 1475
QY 209 sGluArgLeuHisGlnLeuValMetAsn 218
Db GGAACGCTCATCATCAACTTGTATTGAAC 1503
RESULT 8
ID ADQ78551 standard; DNA; 668 BP.
AC ADQ78551;
XX
DT 21-OCT-2004 (first entry)
XX Novel canine microarray-related DNA sequence SeqID101110.
DE
XX canine microarray; drug screening; toxicity assay;
KW environmental pollutant; cellular response; gene expression profile;
KW toxic response; liver necrosis; fatty liver disease;
KW protein adduct formation; hepatitis; dog; ds.
XX Canis familiaris.
OS
XX WO2004063324-A2.
PN
XX 29-JUL-2004.
PD
XX
XX 05-MAY-2003; 2003WO-US013853.
XX
XX 03-MAY-2002; 2002US-0377240P.
PR
XX (GENE-) GENE LOGIC INC.
PA (PFIZ) PFIZER PROD INC.
XX
PI Diggans JC, Potter M, Wei T;
DR WPI; 2004-561890/54.
XX
XX New isolated nucleic acid molecule, useful for drug screening and
PT toxicity assays or for assessing the impact, including toxicity, of a
PT compound, pharmaceutical agent or environmental pollutant on a cell or
PT living organism.
XX
PS Claim 1; SEQ ID NO 10110; 41bp; English.
XX
XX This invention is related to a novel isolated canine nucleic acid
CC sequences and the construction of canine microarrays containing a
CC significant portion of the canine genome. The isolated canine nucleic
CC acid sequences of the invention may be useful for drug screening and
CC toxicity assays. The invention is therefore useful for assessing the
CC impact, including toxicity, of a compound, pharmaceutical agent or
CC environmental pollutant on a cell or living organism. The methods are
CC useful for detecting genes that are up- or down-regulated in canines in a
CC disease state. The sequences are useful as diagnostic agents or markers
CC to detect a cellular response in a sample individually or as part of a
CC gene expression profile. It is also useful as a target for agents that
CC modulate gene expression or activity. The database is useful for
CC producing electronic Northernblots that allow the user to determine the cell
CC type or tissue in which a given gene is expressed and to allow
CC determination of the abundance or expression level of a given gene in a
CC particular tissue or cell. The methods are useful for determining the
CC similarity of a toxic response to one or more individual compounds. The
CC methods are useful for predicting at least one toxic response or the
CC likelihood that a compound or test agent will induce various specific
CC pathologies such as those of the liver (liver necrosis, fatty liver
CC disease, protein adduct formation or hepatitis), those of the kidney,
CC heart, brain or testes, or other pathologies associated with at least one
CC of the toxins. The methods are also useful for predicting or elucidating
CC the potential cellular pathways influenced, induced or modulated by the
CC compound or test agent due to the similarity of the expression profile
CC compared to the profile induced by a known toxin. The present sequence is

CC that of a canine DNA sequence which was claimed for use during the
CC production of a canine microarray of the invention. Note: The sequence
CC data for this patent does not form part of the printed specification but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 668 BP; 205 A; 148 C; 141 G; 157 T; 0 U; 17 Other;

Alignment Scores:
Pred. No.: 3,93e-59 Length: 668
Score: 854.00 Matches: 160
Percent Similarity: 86.3% Conservative: 17
Best Local Similarity: 78.0% Mismatches: 28
Query Match: 27.0% Indels: 0
DB: 13 Gaps: 0

US-10-644-084-2 (1-615) x ADQ78551 (1-668)

QY 411 ThrThrCysAspAspThrThrSerLeuLeuArgAspCysThrLeuLeuGluGluLys 430
DB 1 ACAGCATGCGATGATGACANACGCTCTTCTCGGAGACTGTACTTGCNNNNNAGAA 60
QY 431 GluArgLeuLysGluGluTrpThrLeuPheLysGluLysGluLysAsnPheGluArgGlu 450
DB 61 NNGNNCTCAANNAGATGTTNCCTGTTTAAAGACAAAAGAAATTCGAGAGGAA 120
QY 451 ArgArgSerPheThrGluAlaAlaLeuArgLeuGluLysGluArgLysAlaPheGluGlu 470
DB 121 AGACGAAGCTTTACAGAGAGCCATTCGCTAGGATTCGAGAGAAAGCGTTTGAAGAA 180
QY 471 GluArgAlaSerTrpValLysGlnPheLeuAsnMetThrAsnPheAspHisGlnAsn 490
DB 181 GAAAGAGCCAGTGTGTTAAAGCAACAGATTTTAAACATGACTACCTTTGACCACCAAGAAC 240
QY 491 SerGluAsnValLysLeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIle 510
DB 241 TCAGAAAATATCAAACTTTTCAGTGCTTCTCAGGAAGTTCGTGCGGACAGTCTCTCA 300
QY 511 ValHisSerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCys 530
DB 301 GTGCACCTCAGGCGCAGCAAGAGCCCTCAGTGTCTTAACGGGACTCCAGTTTGC 360
QY 531 ThrSerLysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArgGlnThr 550
DB 361 ACCTCTTAACCTTACTAAGTCTCTTCTGCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 420
QY 551 HisSerCysValSerGluHisSerSerValLeuAsnIleThrProGluGluSer 570
DB 421 CGGTCTCGGCATCTGAGCATAGTTCATCATGACTGAATATTAATCTCTGAAGAACT 480
QY 571 LysProSerGluValAlaArgGluSerThrAspGlnLysTrpSerValGlnSerArgPro 590
DB 481 AAACCAAAATCAGTTCGCGCAGAGAGAGTACAAATCAGAAATGAGCATGCGCATCAAGACT 540
QY 591 SerSerArgGluGlyCysThrSerGlyCysSerSerAlaPheArgSerAlaHisGlyAsp 610
DB 541 GGATCAGCAGAGAGGTGTACAGTGGATGCTCTTTCGACCTACACAAACTCCCATGTAGAA 600
QY 611 ArgAspAspLeuPro 615
DB 601 AAAGATGACTTACT 615

RESULT 9

AAH08911/c
ID AAH08911 standard; cDNA; 563 BP.
XX
AC AAH08911;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (3'-primer) SEQ ID NO:5746.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
XX Claim 3; SEQ ID NO 5746; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX
SQ Sequence 563 BP; 144 A; 115 C; 97 G; 199 T; 0 U; 8 Other;

Alignment Scores:
Pred. No.: 4,28e-48 Length: 563
Score: 715.00 Matches: 145
Percent Similarity: 94.9% Conservative: 4
Best Local Similarity: 92.4% Mismatches: 5
Query Match: 22.6% Indels: 3
DB: 4 Gaps: 1

US-10-644-084-2 (1-615) x AAH08911 (1-563)

QY 63 GluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSerLeuTyr 82
DB 463 GAANCAATTTCTCTCTAC-----CAGGAATGACTACTTTTGGTTTCTTCATTATAT 410
QY 83 GluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeuAsnCys 102
DB 409 GAAGAATCCAAAGGTAAAGACACAAAGAGAGATTAAATATAGTAGTACTTAATTTGT 350
QY 103 MetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerValGluThr 122

Db 349 ATGATGAGCTGCTTGTCTTCAGCGGAAGAACCTTCTAGCTCAGAAAATGTGGAGACA 290
 QY 123 GlnAenLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyAlaLysLeuLys 142
 Db 289 CAGAAATTTGAGCTGGGAAGTGATATGAGACCATCTACAGAGCTGCTACTCAAAACTTAAG 230
 QY 143 GluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArgGlnLeu 162
 Db 229 GAACAACCTGGAAACCTCCAGGAGGGAATGATTGGGCTTCAGGAAGAAGACAGACAGTTA 170
 QY 163 GlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluValGln-Ly 182
 Db 169 CAATGTAGACAGAGAATTGTCATCAGCTACTAAGATGAGAAAGATGAGGTGCAAAA 110
 QY 182 sleuGlnAenIleIleAlaSerArgAlaThrGlnTyrAenHisAspValLysArgLysG1 202
 Db 109 ATTACAAAATATCATTTGCAAGTCGAGCTACTCAGTATATATCATGATATGAAAGAGAAAAGA 50
 QY 202 uArgGluTyrAenLysLeuLysGluArgGluHisGlnLeuValMetAen 218
 Db 49 GCGTGAATATATAAACTGAAGGAACGCTACATCAACTTGTATTGAAC 1

RESULT 10

AAV86670

ID AAV86670 standard; cDNA; 498 BP.

AC AAV86670;

XX 27-APR-1999 (first entry)

DE EST clone BG250.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;

KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;

KW chemotaxis; chemokines; haemostasis; gene therapy; thrombolysis;

KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX Homo sapiens.

XX WO9845435-A2.

XX 15-OCT-1998.

XX 10-APR-1998; 98WO-US006954.

XX 10-APR-1997; 97US-00835913.

XX (GENY) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;

PI Spaulding V, Agostoino MJ;

XX WPI; 1999-070076/06.

XX New polynucleotides encoding human secreted proteins - derived from e.g.

PT human blood, kidney, foetal lung, placenta, testes, brain, ovary,

PT pituitary, retina and colon cDNA libraries.

XX Claim 1; Page 311; 633pp; English.

XX This sequence represents an expressed sequence tag (EST), and is a

CC polynucleotide of the invention. The polynucleotides of the invention are

CC all secreted EST sequences isolated from a variety of human tissue

CC sources. The EST sequences and proteins encoded by them are predicted to

CC have useful biological activities which would make them suitable for

CC treating, preventing or ameliorating medical conditions in humans and

CC animals, although no supporting data is given. Suggested activities

CC include nutritional activity, immune stimulating or suppressing activity,

CC haematopoiesis regulating activity, tissue growth activity,

CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition

CC activity. The EST sequences are also stated to be useful for gene therapy
 XX Sequence 498 BP; 186 A; 71 C; 120 G; 115 T; 0 U; 6 Other;
 SQ

Alignment Scores:

Pred. No.: 1,11e-44 Length: 498
 Score: 671.50 Matches: 136
 Percent Similarity: 90.1% Conservative: 10
 Best Local Similarity: 84.0% Mismatches: 15
 Query Match: 21.2% Indels: 2
 DB: 2 Gaps: 1

US-10-644-084-2 (1-615) x AAV86670 (1-498)

QY 224 IleAlaMetAspValLeuAenTyrValGlyArgAlaAspGlyLysArgGlySerTrpArg 243
 Db 16 CTGCTGCTAGGACATTTTGAATTTATGTCGGGAGAGCTGATGGANAAGAGGCTCCNGGAGG 75
 QY 244 ThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAenAspTyr 263
 Db 76 ACTGGTTAACTGAAGCCNGGAATGA-GATGAATGTATATAAATTTCTTGAATGATTAT 134
 QY 264 GluTyrArgGlnLysGlnIleLeuMetGluAenAlaGluLeuLysLysValLeuGlnGln 283
 Db 135 GAATATCGTCAGAAACAAATCNTAATGAAAAATGCAGAACTTAAGAGGTTCTTCAACAA 194
 QY 284 MetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAla 303
 Db 195 ATGAAAAAGGAATGATTTTCTCTTCTCCCAAAAGAAACCTCGAGAAAGATGA 254
 QY 304 GluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeuSer 323
 Db 255 GATGATAGTACAGGACTGTT--ATTTCGATGTTGAAGAAGATGCCCGGGAACCTAAGC 311
 QY 324 ArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAenSerIle 343
 Db 312 AGAGAGAGTATGTGGACCTTTCTCTGTGAAACTGTGAGAGAGCAGCTTACAAACAGCATC 371
 QY 344 ArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAenGlnAlaSerLys 363
 Db 372 AGAAAAACAGTGGAGAAATTTTGAAGAAGTCATGTAGAAAAGCTTGATAACCAAGTTTCAAG 431
 QY 364 ValHisSerGluGlyLeuAenGluGluAspValIleSerArgGlnAenSPHisGluGlnGlu 383
 Db 432 GTACACCTGGAGAGGTTTTTAATGATGAAGATGTATCTACCGCAAGACCACATGAACAAGAA 491
 QY 384 ThrGlu 385
 Db 492 ACTGAA 497

RESULT 11

AAF66702

ID AAF66702 standard; cDNA; 403 BP.

XX AAF66702;

AC AAF66702;

XX 09-APR-2001 (first entry)

XX Novel human polynucleotide, SEQ ID NO: 2458.

XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;

XX breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

XX WO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-US018374.

XX 02-JUL-1999; 99US-0142310P.

XX 02-JUL-1999; 99US-0142311P.

Db 121 ACAGGAAGCTGTT--ATTTCGATGTTGAAGAAGATGCCGGGAACTAAGCAGAGAGAGT 177
 QY 327 ValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGln 346
 Db 178 ATGTGGGACCTTCTCTGTGAAACTGTGAGAGAGAGCTTTACAAACAGATCAGAAACAG 237
 QY 347 TrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSer 366
 Db 238 TGGAGAAATTTTCAAAAGTCATGTAGAAAGCTTGATAACCAAGTTTCAAAAGGTACACCTG 297
 QY 367 GluGlyLeuAsnGluGluAspValIleSerArgGlnAspHisGluGlnGluThrGluLys 386
 Db 298 GAAGGTTTTAATGTAAGATGTAATCTCACACAGACCATGAAACAAGAACTGAAAAA 357
 QY 387 LeuGluLeuGluIleGluArgCysLysGluMetIleLysAlaGlnGlnLeuLeuGln 406
 Db 358 CTCGAGTTAGAAATTCAGCAGGTGTAAGAAGATGATTAATAACTCAGCAACAGCTTTTACAG 417
 QY 407 GlnGlnLeu 409
 Db 418 CAGCAGCTC 426
 RESULT 13
 ADX36698
 ID ADX36698 standard; cDNA; 1691 BP.
 AC ADX36698;
 XX
 DT 21-APR-2005 (first entry)
 XX
 DE Plant full length insert polynucleotide seqid 19518.
 XX
 KW plant protectant; plant growth regulant; gene therapy; plant;
 KW recombinant DNA construct; physical array; plant breeding marker;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;
 KW growth rate; cell cycle pathway; disease resistance;
 KW galactomannan production; lignin production; plant growth regulator;
 KW yield; plant growth; plant development; seed oil; protein yield;
 KW protein content; gene; ss.
 XX
 OS Unidentified.
 XX
 PN US2004034888-A1.
 XX
 PD 19-FEB-2004.
 XX
 XX 28-APR-2003; 2003US-00425114.
 PF
 PR 06-MAY-1999; 99US-00304517.
 PR 05-NOV-2001; 2001US-00985678.
 XX
 XX (LIU/) LIU J.
 PA (ZHOU/) ZHOU Y.
 PA (KOVA/) KOVALIC D K.
 PA (SCRE/) SCREEN S E.
 PA (TAB/) TABASKA J E.
 PA (CAO/) CAO Y.
 XX
 PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
 XX
 DR WPI; 2004-180133/17.
 XX
 PT New recombinant DNA construct, useful for improving plant tolerance to
 PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
 PT pests, for conferring increased resistance to plant disease, or for
 PT improving yield.
 XX
 PS Claim 1; SEQ ID NO 19518; 15pp; English.
 XX
 CC The invention describes a recombinant DNA construct comprising a
 CC polynucleotide consisting of a sequence encoding an amino acid sequence
 CC available in electronic form from the US patent office at

ftp.segdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide
 of the invention are also useful in physical arrays of molecules and as
 plant breeding markers. The recombinant DNA construct is useful for
 improving plant tolerance to cold, heat, drought, herbicides, extreme
 osmotic conditions, pathogens or pests, for manipulating growth rate in
 plant cells by modification of the cell cycle pathway, for conferring
 lignin or plant growth regulators, for producing galactomannan,
 recombination in plants, for improving yield by modification of
 photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
 or by providing improved plant growth and development under at least one
 stress condition or for modifying seed oil or protein yield and/or
 content. This sequence represents a plant full length insert
 polynucleotide that can be used in the recombinant DNA construct of the
 invention.

Sequence 1691 BP; 474 A; 409 C; 411 G; 397 T; 0 U; 0 Other;

Alignment Scores:	1.08e-20	Length:	1691
Pred. No.:	380.50	Matches:	110
Score:	48.4%	Conservative:	106
Percent Similarity:	24.7%	Mismatches:	155
Best Local Similarity:	12.0%	Indels:	75
Query Match:	13	Gaps:	14

US-10-644-084-2 (1-615) x ADX36698 (1-1691)

QY 27 LysMetSerProSerSerLeuTyr-----SerGlnValLeuCysSer 41
 Db 160 CGGATGTCCTCGTGGCGCGTTCGATCTCGGGGCTCTCGCAGCAGTTG-----CAG 213
 QY 42 SerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGluAsn 61
 Db 214 CGGCACCGTCGCGCCGCGCATGACGACGACGCGGGCGCGCTTCGCAACCGCGGAGAAC 273
 QY 62 IleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSerLeu 81
 Db 274 CTGGAGCACTCGCGCGGTACTCTGAACGACGCTCGTCTCGATTTCCGGATTTCCCGCA--- 330
 QY 82 TyrGluGlnSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeuAsn 101
 Db 331 -----TCTCTCGACCTCTTCCGCCACCGATCCGGTCTCGATTGCAAGAACTTGTAAAC 381
 QY 102 CysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnUserValGlu 121
 Db 382 TGCATCTATGCTGCTGCTCAGCAGCGCAACAAAGGACATTGAGTTCCAGGAGTCTACAAAT 441
 QY 122 ThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLysLeu 141
 Db 442 GACCAGCGCGCAGCGGATCTGATATCTCATCGGCTGGAAGCTTAAGATTGAGAGAAATG 501
 QY 142 LysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArgGln 161
 Db 502 GATGCTCAATTAGCAGCGCAAGATCGAGAGCTAGCCACTCTTACTCGAAGCGGAGCCAAA 561
 QY 162 LeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluValGln 181
 Db 562 AACACTGCTGCTCAAGTCTCAGATTGATTAAGCTGCAACAAAGACGTGATGAATTTTCAG 621
 QY 182 LysLeuGlnAsnIleIleAlaSerArg-----AlaThrGlnTyrAsnHisAspValLys 199
 Db 622 AAAATG-----GTTATAGGAAATCAGCAAGTACGCACCCAGCAAAATTCATGAAATGAAG 675
 QY 200 ArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLys 219
 Db 676 AAGAAAGAAAGAAATACATAAGTTGCAGAAAGCTAAATCATCAGTATTGATGGAGAAAG 735
 QY 220 LysAspLysAsn-----IleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGly 237
 Db 736 AAAAAGGAATCATCCCGTTCTGGAATGGAGATAATGAACCTGTTACAGAAAGAACGACGA 795
 QY 238 LysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLys 257

Qy	139	alysLeuysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAs	159	XX	06-SEP-2000.	
Db	410	AAAGATGGATGCTCAATAGCTGCAAAAGATCGCAGCTGCCACATTGACTAGAACGGA	469	XX	25-FEB-2000; 2000EP-00301439.	
Qy	159	parGlnLeuGlnCysLeuAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGl	179	XX	25-FEB-1999; 99US-0121825P.	
Db	470	GGCCAAAAACACTGCAACTTTGAAAGCTCAGATTCAACAGCTGCAACAGGAACGGCATGA	529	PR	05-MAR-1999; 99US-0123180P.	
Qy	179	uValGlnLysLeuGlnAsnIleIleAlaSerArg-----AlaThrGlnTyrAsnHisAs	197	PR	09-MAR-1999; 99US-0125788P.	
Db	530	GTTCAAAAAATG-----GTTATTGGCAATCAGCAAGTACGTACCCAGCAAAATTCATGA	583	PR	23-MAR-1999; 99US-0126264P.	
Qy	197	pValysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValme	217	PR	29-MAR-1999; 99US-0126785P.	
Db	584	AATGAAGAAAAAGAGAAAGAAATACATCAATTCGAGGAGAAGTTAAACCCAGGTATTAAT	643	PR	01-APR-1999; 99US-0127462P.	
Qy	217	tAsnLysLysAspLysAsn-----IleAlaMetAspValLeuAsnTyrValGlyArgAl	235	PR	08-APR-1999; 99US-0128234P.	
Db	644	CGAGAAAAAGAGGAATCATACGTTTCAGGAATGGAATAATGAACCTTGTTCAGAAAAGA	703	PR	16-APR-1999; 99US-0128714P.	
Qy	235	aAspGlyLysArgGlySerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMe	255	PR	19-APR-1999; 99US-0130077P.	
Db	704	AGCAGCGCAGCTGGACTGTGAGTGGAAAAAAG-----AATGACAATGATTA	751	PR	21-APR-1999; 99US-0130449P.	
Qy	255	tTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAl	275	PR	23-APR-1999; 99US-0130510P.	
Db	752	TTACAAAATGATTGTTGATGCTATGAGTAAAGAACAGAGATTGATGCAAGAAAAATGC	811	PR	28-APR-1999; 99US-0130891P.	
Qy	275	aglLeuLysLysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGl	295	PR	30-APR-1999; 99US-0132048P.	
Db	812	TGATTTCGAGCACTATTGCGTTCGATGCAGATGATATGCGCGAGTTCCTCAATGCTCC	871	PR	04-MAY-1999; 99US-0132484P.	
Qy	295	nLysLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspil	315	PR	05-MAY-1999; 99US-0132485P.	
Db	872	AAATGGAGTTCCA---CACCGAGCTGTTGTGGCAATGACACAGGAGCGAGATCTCC	928	PR	06-MAY-1999; 99US-0132486P.	
Qy	315	eGluAspSerSerglyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrVa	335	PR	07-MAY-1999; 99US-0132863P.	
Db	929	TCAGTCTCCACTTGGT-----GGCAAGACGGATGCTTTGATTTGCCCTTTCACATGGC	982	PR	11-MAY-1999; 99US-0134256P.	
Qy	335	largGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGl	355	PR	14-MAY-1999; 99US-0134218P.	
Db	983	TAGACACCAAGATTGAAGAGAGTTTCGCACATAAATGGCTTCATAAAGGCCCGAATGAC	1042	PR	14-MAY-1999; 99US-0134219P.	
Qy	355	uLysLeuAspAsn-----GlnAlaSerLysValHisSerGluGlyLeuAsnGluLys	373	PR	18-MAY-1999; 99US-0134768P.	
Db	1043	ACAACCTTCAGATGCCCAAGGGTGCAGAGTGCATCAGAAAGCTACC-----	1091	PR	19-MAY-1999; 99US-0134941P.	
Qy	373	pValIleSerArgGlnAepHisGlnGlnGluThrGluLysLeuGluLeuGluLeuGluAr	393	PR	20-MAY-1999; 99US-0135124P.	
Db	1092	-----GAGCGGGAACCTTCAA---CTTGAAGCTCAACTAGTTGA	1126	PR	21-MAY-1999; 99US-0135333P.	
Qy	393	gCysLysGluMetIleLysAlaGlnGlnLeuGlnGln 407		PR	24-MAY-1999; 99US-0135629P.	
Db	1127	AGCTAGAAGCATTATCCAGGACGAGGCATCCCTCATGTGTCGAAG 1169		PR	25-MAY-1999; 99US-0136021P.	
RESULT 15						
AAC43689						
ID	AAC43689 standard; DNA; 1544 BP.					
XX	AAC43689;					
AC	AAC43689;					
XX	18-OCT-2000 (first entry)					
DT	Zea mays DNA fragment SEQ ID NO: 40143.					
DE	Zea mays					
XX	Hybridisation assay; genetic mapping; gene expression control;					
KW	protein identification; signal transduction pathway; metabolic; pathway;					
KW	promoter; termination sequence; corn; ss.					
XX	Zea mays subsp. mays.					
OS	EP1033405-A2.					
XX						
PN						

```
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142777P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.

PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157863P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 1,548-18 Length: 1544
Score: 353.00 Matches: 103
Percent Similarity: 48.6% Conservative: 98
Best Local Similarity: 24.9% Mismatches: 145
Query Match: 11.2% Indels: 69
DB: 3 Gaps: 12

US-10-644-084-2 (1-615) x AAC43689 (1-1544)

Qy 54 GlyValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeu 73
Db 185 GCGCGGTTCCGCAACACGCGGAGAACTGGAGCACTCGCCCGGTACTGTAACACGAGCGCTC 244
Qy 74 ThrThrPheGlyPheProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGlu 93
Db 245 CTCACCTTCGGATTCCCGGCA-----TCTCTCGACCTCTTCGCCACCGATCCG 292
Qy 94 LeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsn 113
Db 293 GTCTCGATTGCAAGAACTTGTAACTGATCTATGCGCTGTGTCAGCAGSANCAGAGGAC 352
Qy 114 LeuLeuAlaGlnGluSerValGluThrGlnAsnLeuLysLeuGlySerAspMetAspHis 133
Db 353 ATTGAGTTTCAGGGAGTCTACAAATGACCAGCGGCA-SGNATGCAATCTGATATCTCACGG 411
Qy 134 LeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGluThrSerArgArgGluMetIle 153
Db 412 CTGGAAGCTAAGATAGAGAGAAATGGATGCTCAATTAGCAGCGAAAGATCGAGAGTAGCC 471
Qy 154 GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu 173
Db 472 ACTCTTACTCGAACSGAGGCCCAACACTGCTGCTCTGTAAGTCTCAGATTGATAAGCTG 531
```

```
QY 174 LysAsnGluLysAspGluValGlnLysLeuGlnAsnIleIleAlaSerArg-----Ala 191
   ::  |||::|||  |||  |||  ::  |||  ::  |||  ::  |||
Db 532 CAACAAGACGATGATTCAGTAAATG-----GTTATAGAAATCAGCAAGTACGC 585
   ::  |||::|||  |||  |||  ::  |||  ::  |||  ::  |||

QY 192 ThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArg 211
   ::  |||::|||  |||  |||  ::  |||  ::  |||  ::  |||
Db 586 ACCCAGCAAAATTCATGAAATGAGAAAGAAAGAAAGAAATACATAAAGTTGCAGGAAAAG 645
   ::  |||::|||  |||  |||  ::  |||  ::  |||  ::  |||

QY 212 LeuHisGlnLeuValMetAsnLysLysAspLysAsn-----IleAlaMetAspValLeu 229
   |||::|||  |||  |||  |||  |||  ::  |||  ::  |||  ::  |||
Db 646 CTAAATCAGGATTGTTGAGAGAGAAAGAAAGAAATCATCCGTTCTGGAATGGAGATAATG 705
   |||  ::  |||  ::  |||  ::  |||  ::  |||  ::  |||

QY 230 AsnTyrValGlyArgAlaAspGlyLysArgGlySerTyrArgThrAspLysThrGluAla 249
   |||  ::  |||  ::  |||  |||  |||  |||  |||  |||
Db 706 AACTTGTTCAGAAAGAACGACCAACCGCGNACATGGAATGGAAAAAAG-----756
   |||  ::  |||  ::  |||  |||  |||  |||  |||

QY 250 ArgAsnGluAspGluMetTyrIleLeuLeuAsnAspTyrGluTyrArgGlnLysGln 269
   |||::|||  |||  |||  |||  |||  |||  |||  |||
Db 757 ---AACGACAATGATTATTACAAAATGATTGTTGATGCATATGAAGTAAAGAAACAAGAG 813
   |||  ::  |||  ::  |||  |||  |||  |||  |||

QY 270 IleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIle 289
   ::  |||  |||  |||  |||  |||  |||  |||  |||
Db 814 CTGATCGAAGAGAAATCGAGATTACGGGCACCTGTTACGTTCAATGCAGATGGACATGCGT 873
   |||  ::  |||  ::  |||  |||  |||  |||  |||

QY 290 SerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGlyThrGlyThr 309
   |||  ::  |||  ::  |||  |||  |||  |||  |||
Db 874 GACTTTCTCAAT-----GCCCGCAATGGGTCAACACAA 906
   |||  ::  |||  ::  |||  |||  |||  |||  |||

QY 310 ValAlaIleSerAspIle---GluAspAspSerGlyGluLeu-----Ser 323
   ::  |||  |||  |||  |||  |||  |||  |||
Db 907 TCTACTGTGTACTGATAATGGAAGACAGAGAGTCAAGGCTCTCTCAGTCTCCGCTTGGTGGC 966
   |||  ::  |||  ::  |||  |||  |||  |||  |||

QY 324 ArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIle 343
   ::  |||  |||  |||  |||  |||  |||  |||
Db 967 AGACCGGATGTTTTGACTTGCCCTTTCATATGCGCCAGAGATCAGATAGAGAGAGTTTG 1026
   |||  ::  |||  ::  |||  |||  |||  |||  |||

QY 344 ArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsn-----GlnAla 361
   |||  ::  |||  |||  |||  |||  |||  |||
Db 1027 CGCACAAAATGACTTCAATCAAGGCACGACATCACACAACACTTCAAGATGCTCAGAAAGGT 1086
   |||  ::  |||  |||  |||  |||  |||  |||

QY 362 SerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHisGlu 381
   ::  |||  |||  |||  |||  |||  |||  |||
Db 1087 GCAGAAAGTGAATCTCTCAGGCAACTGACCGGGAG-----1119
   |||  ::  |||  |||  |||  |||  |||  |||

QY 382 GlnGluThrGluLysLeuGluGluIleGluArgCysLysGluMetIleLysAlaGln 401
   ::  |||  |||  |||  |||  |||  |||  |||
Db 1120 -----CTTGAGCTTGAAGCTCAACTGGTTGAAGCAAGAGCATCATCCAGAGCAG 1170
   |||  ::  |||  |||  |||  |||  |||  |||

QY 402 GlnGlnLeuGlnGlnGlnLeuAlaThr-----411
   ::  |||  |||  |||  |||  |||  |||  |||
Db 1171 GCCTGCATCATGTCCAAGCACTTCAAGTCTGATAAGCCAAAGTGGCAGGAGGCATAGCGGT 1230
   |||  ::  |||  |||  |||  |||  |||  |||

QY 412 -----ThrCysAspAspAspThr 417
   |||  ::  |||  |||  |||  |||  |||  |||
Db 1231 CTGGATGGCGAGCGTGAGGCATCTGCCGAGGTGTAAAAGCATGCTCTGTAGCAAGGAT---1287
   |||  ::  |||  |||  |||  |||  |||  |||

QY 418 ThrSerLeuLeuArgAspCysTyrLeuLeuGluLysGlu 431
   |||::|||  |||  |||  |||  |||  |||  |||
Db 1288 ATATCTGTTCTCGAGGACCTCTCGATTTTAGTAAGTACCGAG 1329
   |||  ::  |||  |||  |||  |||  |||  |||
```

Search completed: June 13, 2006, 23:28:59
Job time : 918 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 13, 2006, 23:33:20 ; Search time 1957 Seconds

(without alignments)
5792.204 Million cell updates/sec

Title: US-10-644-084-2

Perfect score: 3165

Sequence: 1 MGDWMTVPVLCNTENKNLS.....CYSGSSAFPSAHGRDDLP 615

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delpop 6.0	Delpext 7.0	

Searched: 18892170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlh
-Q=/abs/ABSWEB/spool/US10644084/runat_12062006_150429_1550/app_query.fasta_1
-DB=Published Applications NA Main -OFMT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs608
-USER=US10644084 @CEN 1.1 1675 @runat_12062006_150429_1550 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA Main:

```
1: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_PUBCOMB.seq.*
3: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
4: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
5: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
6: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
7: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
8: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
9: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
10: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
11: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
12: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
13: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
14: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11B_PUBCOMB.seq.*
15: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11C_PUBCOMB.seq.*
16: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11D_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2720	85.9	2272	7 US-10-104-047-1497

2	2720	85.9	2272	16	US-11-072-512-1497	Sequence 1497, Ap
3	650	20.5	403	10	US-10-779-543-12377	Sequence 12377, A
4	438	13.8	161874	15	US-11-121-086-75	Sequence 75, Appl
5	380.5	12.0	1691	8	US-10-425-114-19518	Sequence 19518, A
6	379	12.0	1567	8	US-10-424-599-27277	Sequence 27277, A
7	378	11.9	1805	9	US-10-425-115-157154	Sequence 157154, A
8	361.5	11.4	3750	8	US-10-437-963-34854	Sequence 34854, A
9	334.5	10.6	1062	9	US-10-767-795-5235	Sequence 5235, Ap
10	338	10.4	513	7	US-10-029-386-5594	Sequence 5594, Ap
11	290	9.2	1199	8	US-10-425-114-10108	Sequence 10108, A
12	285	9.0	178	7	US-10-029-386-19350	Sequence 19350, A
13	234.5	7.4	5040	7	US-10-369-493-45767	Sequence 45767, A
14	233	7.4	730	8	US-10-767-701-7320	Sequence 7687, A
15	233	7.4	3444	8	US-10-437-963-77055	Sequence 77055, A
16	231.5	7.3	5373	11	US-10-932-182A-5516	Sequence 5516, Ap
17	230.5	7.1	5040	9	US-10-828-985A-10	Sequence 10, Appl
18	225.5	7.1	7509	9	US-09-960-253-145	Sequence 145, App
19	223.5	7.1	10300	3	US-09-960-253-145	Sequence 49, Appl
20	223.5	7.1	10300	9	US-10-422-524-49	Sequence 122, App
21	223.5	7.1	10300	10	US-10-947-249-122	Sequence 45972, A
22	220.5	7.0	6386	6	US-10-369-493-45372	Sequence 40, Appl
23	220.5	7.0	6386	6	US-10-098-841-40	Sequence 40754, A
24	219.5	6.9	2602	13	US-11-097-143-40754	Sequence 134, App
25	218	6.9	11677	6	US-10-082-830-134	Sequence 6, Appl1
26	217.5	6.9	4470	7	US-10-205-219-6	Sequence 117, App
27	217	6.9	3614	13	US-11-097-143-33485	Sequence 117, App
28	217	6.9	6921	3	US-09-735-705-117	Sequence 117, App
29	217	6.9	6921	3	US-09-850-716A-117	Sequence 117, App
30	217	6.9	6921	3	US-09-897-778-117	Sequence 117, App
31	217	6.9	6921	3	US-09-466-368A-117	Sequence 117, App
32	217	6.9	6921	6	US-10-007-700-117	Sequence 117, App
33	217	6.9	6921	7	US-10-117-982-117	Sequence 117, App
34	217	6.9	6921	9	US-10-313-986-117	Sequence 117, App
35	217	6.9	6921	9	US-10-775-972-117	Sequence 117, App
36	217	6.9	6921	10	US-10-922-124-117	Sequence 117, App
37	217	6.9	6921	10	US-10-623-155-117	Sequence 28, Appl
38	217	6.9	7045	3	US-09-919-172-28	Sequence 28, Appl
39	217	6.9	7045	9	US-10-752-986-28	Sequence 170, App
40	217	6.9	8684	7	US-10-205-219-170	Sequence 223, App
41	217	6.9	8684	10	US-10-756-149-223	Sequence 649, App
42	217	6.9	8684	10	US-10-287-436A-649	Sequence 219, App
43	217	6.9	8974	7	US-10-341-434-219	Sequence 229, App
44	217	6.9	8974	7	US-10-341-434-229	Sequence 17, Appl
45	217	6.9	8974	10	US-10-287-436A-17	

ALIGNMENTS

RESULT 1
US-10-104-047-1497
; Sequence 1497, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1497
; LENGTH: 2272
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1497

Alignment Scores:	Pred. No.:	Length:
Score:	5.3e-224	2272
Percent Similarity:	2720.00	Matches: 536
Best Local Similarity:	86.8%	Conservative: 34
	81.6%	Mismatches: 40

Db 1769 GTTCTTAATATGACTACCTTTGACCACCAGAACTCAGAAAATGTGAACCTTTTCAGTGC 1828
Qy 499 aPeSerGlySerSerAspProAspAenLeuIleValHisSerArgProArgGlnIleVal 519
Db 1829 CTTCTCAGGAAGTTCGATTTGGACAACTTTATAGTGCACTCGAGCAGCCCGCAAGAA 1888
Qy 519 sLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuPr 539
Db 1889 GCCTCACAGTGTCTAATGAGTCTCCAGTTTGCATGCTCTAACTTACTTAATCTCTTCC 1948
Qy 539 oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe 559
Db 1949 TGCTTCACCTTCACCTTCAGACTTTTGCAGACACAGCTTCTGCATATCTGAACATAGTTC 2008
Qy 559 rIleSerValLeuAenIleThrProGluGluSerLysProSerGluValAlaArgGluSe 579
Db 2009 AATCAATGACTGAATATAACTGCTGAAGAAATTAACCAATCAGTTTGGAGGAGACG 2068
Qy 579 rThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGl 599
Db 2069 TACAATAAATAATGAGTGTGGCGTCAAGACCTGGATCACAGGAAGTTGCTATAGTGG 2128
Qy 599 yCysSerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
Db 2129 ATGCTCTTTCAGCTACACAAATTTCTCATGTAGAAAAAGATGACTTACCT 2177

RESULT 3
US-10-779-543-12377
; Sequence 12377, Application US/10779543
; Publication No. US20050227917A1
; GENERAL INFORMATION:
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
; TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/076,555
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 09/217,471
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: 60/068,755
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/080,664
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/105,234
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 09/297,648
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: PCT/US99/01619
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: 60/072,910
; PRIOR FILING DATE: 1998-01-28
; PRIOR APPLICATION NUMBER: 60/075,954
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/080,114
; PRIOR FILING DATE: 1998-03-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12377
; LENGTH: 403
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-779-543-12377

Alignment Scores:
Pred. No.: 2,02e-46 Length: 403
Score: 650.00 Matches: 124
Percent Similarity: 97.7% Conservatve: 4
Best Local Similarity: 94.7% Mismatches: 3
Query Match: 20.5% Indels: 0
DB: 10 Gaps: 0

US-10-644-084-2 (1-615) x US-10-779-543-12377 (1-403)
Qy 132 AsphHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGluThrSerArgArgGlu 151
Db 10 GACCATCTACAGAGCTGCTACTCAAACTTTATGGAAACAACCTCGAAACCTCCAGAGGGAA 69
Qy 152 MetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGln 171
Db 70 ATGATTGGGCTTCAGGAAAGAGACAGAGTACAAATGTAAAGAACAGGAATTTGCATCAG 129
Qy 172 LeuLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArgAla 191
Db 130 CTACTAAAGAATGAGAAAGATGAGGTGCAAAAATTTACAAAATATCATTTGCAAGTCGAGCT 189
Qy 192 ThrGlnTrpAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArg 211
Db 190 ACTCAGTATAATCATGATATGAAGAGAAAGAGCGTGAATATAATAAATCTGAAGGAACGT 249
Qy 212 LeuHisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyr 231
Db 250 CTACATCAACTTGTATTGACACAGAAAGATAAGAAAAATAGCTATGACACATTTGAATTAT 309
Qy 232 ValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsn 251
Db 310 GTCGGGAGAGCTGATGGAAAAAGAGGCTCTCGGAGGACTGGTTAAAAAATGAAAGCCAGGAAT 369
Qy 252 GluAspGluMetTyrLysIleLeuLeuAsnAsp 262
Db 370 GAAGATGAATGTATATAAAATTTCTTTGAATGAT 402

RESULT 4

US-11-121-086-75/c
; Sequence 75, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 75
; LENGTH: 161874
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-75

Alignment Scores:
Pred. No.: 1.27e-24 Length: 161874
Score: 438.00 Matches: 114
Percent Similarity: 32.0% Conservatve: 10
Best Local Similarity: 29.5% Mismatches: 12
Query Match: 13.8% Indels: 252
DB: 15 Gaps: 1

US-10-644-084-2 (1-615) x US-11-121-086-75 (1-161874)
Qy 10 ProValLeuCysThrGluAsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSer 29
Db 57511 CCTTTTATCTTTCAGAAAGCAAACTATCTCTCAATATACCTCAGAAACAAAGATGTCT 57452
Qy 30 ProSerSerLeuTyrSerGlnGlnValLeuCysSerSerValProLeuSerLysAsnVal 49
Db 57451 CCATCAAGTTTATCTACTCAGCAAGTGTCTTCTTCAATACCTTTATCGAAAAATGTG 57392
Qy 50 HisGlyValPheGlyValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyrLeu 69
Db 57391 CACAGTTTTTTCAGTGCCTTCTGACACAGAAGATAATATTGAACAGATATCTCATATCTT 57332

Db 736 AAAAAGGAATATCCCGTTCTGGAATGGAGATAAATGAACTTGTATCAGAAAGAGACGA 795
QY 238 LysArgGlySerTrpArgThrAspLysThrGluAlaAArgAsnGluAspGluMetTyrLys 257
Db 796 CAACGGGAACATGGAAATGGAATAAG-----AACGCAATGATTATTACAAA 843
QY 258 IleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeu 277
Db 844 ATGATTGTTGATGCATATGAAGTGAAGAAACAAGAGCTGATCAAGAGAAATCGATTTA 903
QY 278 LysLysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLys 297
Db 904 CGGGCACTGTTACGTTCAATGCAGATGGAGATGCGTGCATTTCTCAAT----- 951
QY 298 LysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIle---Glu 316
Db 952 -----GCCCGAATGGGTCAATCACAATCTACTGTTACTGATTAATGGAAGA 996
QY 317 AspAspSerGlyGluLeu-----SerArgAspSerValTrpGlyLeuSer 331
Db 997 CAGGAGTCAGGCTCTCTCAGTCTCGCTTGGTGGCAAGACGGATGTTTTCAGTTGCC 1056
QY 332 CysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLys 351
Db 1057 TTTTCATATGCCAGAGATCAGATAGAAAGAGAGTTTTCGCCACAAAAATGACTTCAATCAAG 1116
QY 352 SerHisValGluLysLeuAspAsn-----GlnAlaSerLysValHisSerGluGlyLeu 369
Db 1117 GCACGAATGACAACTTCAAGATGCTCAGAAAGGTGCAGAAAGTGACTTCTCAGGCAACT 1176
QY 370 AsnGluGluAspValIleSerArgGlnAspHisGluGlnIleuThrGluLysLeuGluLeu 389
Db 1177 GACCGGAG-----CTTGAGTTGAGCT 1200
QY 390 GluIleGluArgCysLysGluMetIleLysAlaGlnGlnLeuGlnGlnGlnLeu 409
Db 1201 CAACGTGTTGAAGCAAGAGCATCATCCAAAGCAGGCGCTGCATCATGTCCAAGCACTTC 1260
QY 410 AlaThr----- 411
Db 1261 AAGTCTGATAAGCAAGTGGCAGGAGGCATACGGTCTGGATGGCGAGCGTGAGGCATCT 1320
QY 412 -----ThrCysAspAspAspThrThrSerLeuLeuArgAspCysTyr 425
Db 1321 GCGAGGTGTAAAGCATGTCTGTAGCAAGGAT---ATATCTGTCTTCGCAAGACCTCCTG 1377
QY 426 LeuLeuGluGluLysGlu 431
Db 1378 ATTTTAGTAGTACCAG 1395

RESULT 6

US-10-424-599-27277
; Sequence 27277, Application US/10424599
; Publication NO. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yinui
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5323)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 27277
; LENGTH: 1567
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_124632C.1
US-10-424-599-27277

Alignment Scores:
Pred. No.: 3, 03e-22 Length: 1567
Score: 279.00 Matches: 94
Percent Similarity: 52.8% Conservative: 95
Best Local Similarity: 26.3% Mismatches: 143
Query Match: 12.0% Indels: 26
DB: 8 Gaps: 8
US-10-644-084-2 (1-615) x US-10-424-599-27277 (1-1567)
QY 56 PheCysThrGlyGluAsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThr 75
Db 176 TTCCCGCATCGGCAAAATTTGGAGCATTCGCGTAAGTATCTCAACCAATCGCTCGTCACG 235
QY 76 PheGlyPheProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsn 95
Db 236 TTTGGTTTCCCGGCT-----TCGTCGATCTCTTCGCAAAATGACCCCGTTTCA 283
QY 96 IleValAlaValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeu 115
Db 284 ATTCCGAGGACTTGCAATTGCAATTACGCTTGTCTGCAGACAGACAGCGCGACGTGGAG 343
QY 116 AlaGlnGluSerValGluThrGlnAsnLeuLysLysLeuGlySerAspMetAspHisLeuGln 135
Db 344 TTTAGAGAGTCTGCTTAATGATCAAAAGACACAGCACTGTTGTCTGACATTTTCGAGATTGGAG 403
QY 136 SerCysTyrAlaLysLeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeu 155
Db 404 GCCAAAGTGGAGAGGCTTGAAGGTCACACTCAAGTCAAAGACAGGGAAGATAGCGACTATT 463
QY 156 GlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsn 175
Db 464 ACTAGACGGAAGCTAAAAACACTGCAGCTCTGAAGGCCCAATTTGAGAAGCTGCAACAG 523
QY 176 GluLysAspGluValGlnLysLeuGlnAsnIleIleAlaSerArg-----AlaThrGln 193
Db 524 GAGAGGATGAATTCACAGAAAGT-----GTTATTGTAATTCAGCAAGTAAAGACTCAA 577
QY 194 TyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHis 213
Db 578 CAATGTCATGATGATGAAGAAAGAAAGAAATACATAAAGTTTCGAGAGAGGCTAAAC 637
QY 214 GlnLeuValMet---AsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrVal 232
Db 638 CAAGTGTGTTGAGAGAAAGAGAGCTAGGTCAGGCATGGAGATAATGAATTTGGCTT 697
QY 233 GlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGlu 252
Db 698 CAGAAGGAAGGAGGAGCAACGTTGGAACATGGAAAGAAAGCT-----GAC 745
QY 253 AspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMet 272
Db 746 AATGATTTTATAAAAGATTGTGGATGCTTATGATCAAAAAAATCAAGAACTGATGCA 805
QY 273 GluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeu 292
Db 806 GAGAATGCTGATTAAAGAGCATTTAGATCAATGCAGGTGGATATCGTGATTCTTA 865
QY 293 SerProGlnLysLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIle 312
Db 866 AATGCTCTTAATGAGTACCAAAAG-----CAATCTTTTCAGTGGCAATGAAAGAGTT 916
QY 313 SerAspIleGluAspAspSerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCys 332
Db 917 GAAAGTGATCTCTCACAATCTCCATTGTTGGGAAAAATGGATGATATTGACCTTCTCTT 976
QY 333 AspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSer 352
Db 977 CACATGGCCAGAGATCAAAATAGAGAAAGTCTTCGCAACAAAAATGGCTTCTATAAAGGAG 1036
QY 353 HisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGluGlyLeuAsnGlu 372
Db 1037 CGTATGGTTCAATTG-----CAAGATGCTCAAAAAGAA 1069

QY 373 AspValIleSerArgGlnAspHisGluGlnGlnThrGluLysLeuGluLeuGluLeuGlu 392
 Db 1070 GCAGAGATTACTTCAGAGGCTACTGAGAGGAACTTGAG---CTTGAAGCTCAACTGTGC 1126
 QY 393 ArgCysLeuMetIleLysAlaGlnGlnGlnLeuGluGlnGlnGlnLeuAla 410
 Db 1127 GAAGCAGGAGCATCATACAGGACGAGCATCAATTAATGTCTAAACATCTTGCA 1180

RESULT 7
 US-10-425-115-157154
 ; Sequence 157154, Application US/10425115
 ; Publication No. US20040214272A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(5322)B
 ; CURRENT APPLICATION NUMBER: US/10/425,115
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 369326
 ; SEQ ID NO 157154
 ; LENGTH: 1805
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: MRT4577_74904C.1
 US-10-425-115-157154

Alignment Scores:
 Pred. No.: 4,47e-22 Length: 1805
 Score: 378.00 Matches: 103
 Percent Similarity: 49.0% Conservative: 100
 Best Local Similarity: 24.9% Mismatches: 143
 Query Match: 11.9% Indels: 68
 DB: 9 Gaps: 12

US-10-644-084-2 (1-615) x US-10-425-115-157154 (1-1805)

QY 54 GlyValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeu 73
 Db 287 GGGCGGTTTCGAAACGCGGAACTTGGAGCACTGCGCCGCTGACCTGAACACAGACGCTC 346
 QY 74 ThrThrPheGlyPheProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGlu 93
 Db 347 GTCACCTTCGGATTCCCGCA-----TCTCTCGACCTCTTCGCCACCGATCCG 394
 QY 94 LeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsn 113
 Db 395 GTCTCGATTGCAAGAACTTGTAACTGCATCTATGCGCTCTGCAGCAGCGACAGAGGGAC 454
 QY 114 LeuLeuAlaGlnGluSerValGlnThrGlnAsnLeuLysLeuLysLeuLysSerAspMetAspHis 133
 Db 455 ATTGAGTTTCAGGAGGCTCAAAATGACACGCGCAGCGATGCAATCTGATATCTCACGG 514
 QY 134 LeuGlnSerCysTyrAlaLysLeuLysGluGlnGluLeuThrSerArgArgGluMetIle 153
 Db 515 CTGGAAGCTTAAGATAGAGATGGATGCTCAATTAGCAGGAAAGATCGAGAGCTAGCC 574
 QY 154 GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu 173
 Db 575 ACTCTTACTCGAAGCGAGCCCAAAACACTGTGCTCTCGAAGTCTCAGATTGATAAGCTG 634
 QY 174 LysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArg-----Ala 191
 Db 635 CAACAGAAACGCTGATGAATTTTCAGAAATG-----GTTATAGAAATCAGCAAGTACGC 688
 QY 192 ThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArg 211
 Db 689 ACCCAGCAATTCATGAAATGAAGAAGAAGAAAGAAATACATAAAGTTGCAGGAAG 748

QY 212 LeuHisGlnLeuValMetAsnLysLysAspLysAsn-----IleAlaMetAspValLeu 229
 Db 749 CTAATTCAGGTATTGATGAGAGAAAGAAAGAAATCATCCGTTCTGGAATGGAGATAATG 808
 QY 230 AsnTyrValGlyArgAlaAspGlyLysArgGlySerTyrArgThrAspLysThrGluAla 249
 Db 809 AACTTGTTCACAGAAAGAGGACGACAAACCGCAACATGGAATGGAAGAAAG----- 859
 QY 250 ArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGln 269
 Db 860 ---AACGACAATGATTATTACAAATGATTGTTGATGTCATATGACATGAGAAACAAGAG 916
 QY 270 IleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIle 289
 Db 917 CTGATGCAAGAGATGCAAGATTACGGGCACCTGTTACGTTCAATGCAATGAGATGCGT 976
 QY 290 SerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGlyThrGlyThr 309
 Db 977 GACTTTCTCAAT-----GCCCGAATGGTCATCACA 1009
 QY 310 ValAlaIleSerAspIle---GluAspAspSerGlyGluLeu-----Ser 323
 Db 1010 TCTACTGTTACTGATAATGGAAGACAGAGTACGGGTCTCCTCAGTCTCGCTTGGTGGC 1069
 QY 324 ArgAspSerValTyrGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIle 343
 Db 1070 AAGACGGATGTTTTGACTTGCCCTTTCATATGGCCAGAGATCAGATAGAGAGATTG 1129
 QY 344 ArgLysGlnTyrArgIleLeuLysSerHisValGluLysLeuAspAsn-----GlnAla 361
 Db 1130 CGCACAAAAATGACTTCAATCAAGCGCAAGATGACACAACTTCAAGATGCTCAGAAAGT 1189
 QY 362 SerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHisGlu 381
 Db 1190 GCAGAAAGTACTTCTGAGCAACTCACCGGAG-----1222
 QY 382 GlnGluThrGluLysLeuGluLeuGluLeuArgCysLysGluMetIleLysAlaGln 401
 Db 1223 -----CTTGAGCTTGAAGCTCAACTGTTGAAGCAAGAGCATCATCAAGAGCAG 1273
 QY 402 GlnGlnLeuLeuGlnGlnLeuAlaThr-----1411
 Db 1274 GCCTGCATCATGCTCCAAAGCACTTCAAGTCTGTATAAGCCAAAGTGGCAGGAGCATACGGT 1333
 QY 412 -----ThrCysAspAspThr 417
 Db 1334 CTGGATGGCAGCGTGAGCATCTCCGAGGTTGTAAGCATGTCTGTAGCAAGGAT--- 1390
 QY 418 ThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLysGlu 431
 Db 1391 ATATCTGTTCTGCAAGACCTCTGATTTTAGTAAGTACCGAG 1432

RESULT 8

US-10-437-963-34854/c
 ; Sequence 34854, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
 ; FILE REFERENCE: 38-21(5322)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 34854

;
; LENGTH: 3750
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_3882C.1
US-10-437-963-34854

Alignment Scores:
Pred. No.: 3 17e-20 Length: 3750
Score: 361.50 Matches: 96
Percent Similarity: 51.8% Conservative: 92
Best Local Similarity: 26.4% Mismatches: 144
Query Match: 11.4% Indels: 31
DB: 8 Gaps: 10

US-10-644-084-2 (1-615) x US-10-437-963-34854 (1-3750)

```
Qy 51 GlyValPheGlyValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyrLeuAsp 70
Db 3561 GCGCGCGCGCGCGTTCGCGAGCGCGGGGAACTCGAGCACTGCGCCAGGTACCTCAAC 3502
Qy 71 GlnGluLeuThrThrPheGlyPheProSerLeuTyrGluGluSerIleSerLysGluAla 90
Db 3501 CAGACGCTGCTACCTTCGGCTTCGCGCC-----TCGCTCGATCTCTTCGCC 3454
Qy 91 LysArgGluLeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuValLeuGln 110
Db 3453 ACGGACCCGATTCGATTCGGAGACGTCGCACTGCATCTATGCTCTTCAGCAGCGG 3394
Qy 111 ArgLysAsnLeuAlaGlnGluSerValGluThrGlnAsnLeuLysLeuGlySerAsp 130
Db 3393 CAGCGTGATATTGAGTTCGGGAGTCTACCAATGATCTCGCTCAGGTATGAGTCTGAT 3334
Qy 131 MetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuThrSerArgArg 150
Db 3333 ATTCTAGATTAGACCCCAAGATTGAAGAATGGATGCTCAATTAGCTGCAAAAGATCGC 3274
Qy 151 GluMetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHis 170
Db 3273 GAGTGGCCACATTGACTAGACGGAGGCCAAACACATGCACTTTGAGGCTCAGATT 3214
Qy 171 GlnLeuLeuLysAsnGluLysAspGluValGlnGlnLysLeuGlnAsnIleIleAlaSerArg 190
Db 3213 GACAAGCTCAACAGGAACGCGATGAGTTCCAAAAAATG-----GTTATTGCAATCAG 3160
Qy 191 -----AlaThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeu 208
Db 3159 CAAGTACGTACCCAGCAAAATTCATGAATGAAGAAAAAGAGAAAAAGAAATACATCAAAATTG 3100
Qy 209 LysGluArgLeuHisGlnLeuValMetAsnLysLysAspLysAsn-----IleAlaMet 226
Db 3099 CAGGAGAGTTAAACAGGATTAATGAGAAAGAAAGAAAGAAATCATCAGTTCAAGATG 3040
Qy 227 AspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLys 246
Db 3039 GAAATAATCAACTGTTGTCAGAAAGAACGCGCGTGAACTTGGAGTGGAAAAAAG 2980
Qy 247 ThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArg 266
Db 2979 -----AATGACAATGATTATTACAAATGATTGTTGATGCTTATGAGGTAAAG 2932
Qy 267 GlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnMetLysLys 286
Db 2931 AAGCAAGATTGATGCAAGAAATGCTGATTTCGAGCACTATTGCTCGATGAGATG 2872
Qy 287 GluMetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGly 306
Db 2871 GATGCGCGAGTTCCTCAATGCTCCAAATGGAGTTCCA---CAGCCAGCTGTTGGGCG 2815
Qy 307 ThrGlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeuSerArgAspSer 326
Db 2814 AATGGACGACAGGAGCGAGGATCTCTCTCAGTCTCCACTTTGGT-----GGCAAGACGGAT 2761
```

```
Qy 327 ValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGln 346
Db 2760 GTCITTGATTGCCCTTTTCACATGCATGAGACACAGATTGAAGAGATTGGCGCACTAAA 2701
Qy 347 TrpArgIleLeuLysSerHisValGluLysLeuAspAsn-----GlnAlaSerLysVal 364
Db 2700 ATGGCTTCATAAAGCGCCGAATGACACAACTTCAAGATGCCAAAGGGTGCAGAAGTG 2641
Qy 365 HisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHisGluGlnGluThr 384
Db 2640 ACATCAGAAAGTACC-----GAGCGGAACATT 2614
Qy 385 GluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAlaGlnGlnLeu 404
Db 2613 GAA---CTTGAAGCTCAACTAGTTCAAGCTAGAACATTATCCAGGAGCAGGCATCCCTC 2557
Qy 405 LeuGlnGln 407
Db 2556 ATGTCCAAG 2548
RESULT 9
US-10-767-795-5235
; Sequence 5235, Application US/10767795
; Publication No. US20040181830A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53534)B
; CURRENT APPLICATION NUMBER: US/10767,795
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 5235
; LENGTH: 1062
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C71981_1
US-10-767-795-5235
Alignment Scores:
Pred. No.: 1 24e-18 Length: 1062
Score: 334.50 Matches: 95
Percent Similarity: 50.8% Conservative: 95
Best Local Similarity: 25.4% Mismatches: 140
Query Match: 10.6% Indels: 44
DB: 9 Gaps: 11
US-10-644-084-2 (1-615) x US-10-767-795-5235 (1-1062)
Qy 20 SerGlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeu 39
Db 3 TCCTCGCTCTCTCTCTCAGACAAA-----TACACACAGTTTCTCTCA 44
Qy 40 CysSer-----SerValProLeuSerLysAsnVal 49
Db 45 ATCTCTCTCGCAATGCCGCACTGATGCAGAAATGGATCTCCCTCTCTCAGTCGCC 104
Qy 50 HisGlyValPheGly---ValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyr 68
Db 105 TTCTCTATCGGAGAGTATACATTTGCAGATGCTGGAACCTTGGACCACTGTACAAAGTAC 164
Qy 69 LeuAspGlnGluLeuThrThrPheGlyPheProSerLeuTyrGluGluSerLysSerLys 88
Db 165 TTGAACACAGACGCTCGTTACCTTTGGCTTCTCTGCT-----TCTCTCGATCTC 212
Qy 89 GluAlaLysArgGluLeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuVal 108
Db 213 TTTGCCAATGATCCGTTTCGGTCGCGAGGACCTGCATTTGCATATATTCATTGCTGAG 272
Qy 109 LeuGlnArgLysAsnLeuLeuAlaGlnGluSerValGluThrGlnAsnLeuLysLeuGly 128
```

```
Db 273 CACGGCGACGCGACATGTAATTTAGAGAGTCTGCTTAATGACCAAGACACACACATTATTA 332
QY 129 SerAspMetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGluThrSer 148
Db 333 TCAGATATCTCAAGATTAGAGCGCAAAAGTTGAGAGGCTTGATGCACAGTTGCAAGCTAAA 392
QY 149 ArgArgGluMetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSer 168
Db 393 GATAGAGAGATGACCAACAATGACTAGAACGGACCAAAAATAACTGCAGCTTTTAAGGCA 452
QY 169 LeuHisGlnLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAla 188
Db 453 CAGATTGAGAGAGCTAGACGACGAGACGAGATGAATTTTCAGAGGATG-----GTGATGGT 506
QY 189 SerArg-----AlaThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsn 206
Db 507 AATCAGCAAGTGAAGAACTCAACAGATACATGAAATGAAAGAAAGAAAGGAGGTACATA 566
QY 207 LysLeuLysGluArgLeuHisGlnLeuValMet---AsnLysLysAspLysAsnIleAla 225
Db 567 AAGTTTCAGAGAGGCTGACCAAGCTCTTAATGCAAAAGAAAGAAAGAAATCAAGATCAGGA 626
QY 226 MetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTyrArgThrAsp 245
Db 627 ATGGAATATGAAATTTGCTTCAGAAAGAGGGCGCAACGTGGACCTGGAAATGGGAAG 686
QY 246 LysThrGluAlaArgAsnGluMetTyrLysIleLeuLeuAsnAspTyrGluTyr 265
Db 687 AAAGCA-----GACAATGATTTCTATAAAAGATTTGCGATGCTTATGAGGCA 734
QY 266 ArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysValLeuGlnGlnMetLys 285
Db 735 AAAAATCAAGAACTGAAGACAGAAATGCTGATTTAAGAGACATTAATGCGCTCAATGCAG 794
QY 286 LysGluMetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAsp 305
Db 795 GTAGATATCGGTGATTTCTTAATGCTCTTAATGGTTGCCAAAGCA-----842
QY 306 GlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGlyLysLeuSer-----323
Db 843 -----TCTTTGCTGTTAGCGAAAGACATGAAATGATCATCACATCGCCACTT 893
QY 324 -----ArgAsp-SerValTyrGlyLeuSerCysAspThrValArgGluGlnLeuThrAs 341
Db 894 GTGGAGGCGGAGATGCTTTGATCTCTTCATCGCTAGGATGATGATGATGATGATGATGATG 953
QY 341 nSerIleArgLysGlnTyrArgIleLeuLysSerHisValGluLysLeuAspAsn-----359
Db 954 AAGTCTTCGACTAAGATGGCTTCTGTAAAGAGCGGATGGTTCAATTACAGGATCGCGCA 1013
QY 360 -GlnAlaSerLysValHisSerGlyGlyLeuAsnGluGlu 372
Db 1014 AAAAGGTGCTGAAGTCACCTCTCTGACCAACACAGAGGGAG 1053

RESULT 10
US-10-029-386-5594/c
; Sequence 5594, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 5594
; LENGTH: 513
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009265.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EST HUMAN HIT: BE694886.1, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: g114723835, EVALUE 3.00e-98
; OTHER INFORMATION: SWISSPROT HIT: Q40554, EVALUE 1.20e+00
US-10-029-386-5594

Alignment Scores:
Pred. No.: 1,69e-18 Length: 513
Score: 328.00 Matches: 76
Percent Similarity: 62.5% Conservative: 4
Best Local Similarity: 59.4% Mismatches: 13
Query Match: 10.4% Indels: 35
DB: 7 Gaps: 2

US-10-644-084-2 (1-615) x US-10-029-386-5594 (1-513)
QY 237 GlyLysArgGlySerTyrArgThrAspLysThrGluAla-----249
Db 511 GGAAGAAAGAGGCTCTCTGGAGGACTGGTAATACTGAAGCCAGGTAAAGTCCAGTTAAATT 452
QY 250 -----250
Db 451 GATAAATGTAATCAGACTTCTTCTATCTCTGCAATTTTCATCTGACAGTTTCTTCTCTCAG 392
QY 250 gheGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIle 270
Db 391 GAATGAAGATGAAATGTATATAAAATTTCTTCTGAATGATTAATGATATCGTCAGAAACAAAT 332
QY 270 eLeuMetGluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIleSe 290
Db 331 CCTAATGGAATGCAGAACTTAAGAGGTTCTTCAACAATGAAAAGAAATGATGTTTC 272
QY 290 rLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGlyThrGlyThrVa 310
Db 271 TCTTCTTCTTCTCCCAAGAAAGAAACCTAGAGAAAGAGTAGATGATGATGATGATGATGATGAT 212
QY 310 lAlaIleSerAspIleGluAspAspSerGlyGluLeu-----322
Db 211 AAGTGGCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 152
QY 323 -SerArgAspSerValTyrGly 329
Db 151 TGGTAGGCATTTCTCAATGGGA 130

RESULT 11
US-10-425-114-10108
; Sequence 10108, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 10108
; LENGTH: 1199
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700894116_FLI
US-10-425-114-10108
```


Alignment Scores:

Pred. No.:	4.28e-09	Length:	5040
Score:	234.50	Matches:	145
Percent Similarity:	36.4%	Conservative:	124
Best Local Similarity:	19.6%	Mismatches:	230
Query Match:	7.4%	Indels:	241
DB:	7	Gaps:	28

US-10-644-084-2 (1-615) x US-10-369-493-45767 (1-5040)

QY	15	GluAsnLysAsnLeuSerGlnTyrThr	-----	23
DB	3028	GAATAATAAGATCTATCTGAATTGGTTATCCGCCCTTGAAAGGATCGCGCTGATGCCAA	3087	
QY	24	SerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCysSerVal	43	
DB	3088	GCAGAAATTACAAAAACCAATCTCTATTGTATAGCGCTCAGGATTTA	3135	
QY	44	ProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu	-----	60
DB	3136	---CTGATTAACATGAAGGAATGGATGGAGAAAGGCTGATTATGAAGAGAAATTG	3192	
QY	61	-----AsnIleGluGlnSer	-----	68
DB	3193	ATATCTAATATTGAACAGCGGAATCTCTTCGCTGTTGAAATTCGGTATTGATTGAAAG	3252	
QY	69	LeuAspGlnGluLeuThrPheGly	-----	78
DB	3253	GTTGATGCACACAGCAGCAATAATATGGATGAAGATCAATTTGAAGCTAGTTCTTTGGTTT	3312	
QY	79	ProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAla	98	
DB	3313	TCAACTTACGTCATCAAGAAATCTTTAGACGCAAG	3351	
QY	99	ValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlu	118	
DB	3352	TTAACCACCTGTAAAGAGAGATTGGCATTCGTTAAACAAAAGAACGATAGTTAGAAAAA	3411	
QY	119	SerVal---GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCys	137	
DB	3412	ACGATAATGACCTGCACAGCAACCCAGACTTTATCGGAGAAGGAATAC	3465	
QY	138	TyrAlaLysLeuLysGluGlnLeuGluThrSerArgArgGluMetIleGly	-----	154
DB	3466	TCGCGAGTTATTATTCAGCAATTCAGAGATATCACGAAGAAGTTACTCAAGTTACATTT	3525	
QY	155	-----LeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArg	167	
DB	3526	TTAABAGAGATAATGCCATCTTACAAAGAGCTCTCAAAAACGTAAACCGAAAAAACAGG	3585	
QY	168	SerLeuHisGlnLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleIle	187	
DB	3586	GAATTTATAACAATTAATATGATAGCAAGAGAAATTTCTCGATTACAGCGTGATCTC	3645	
QY	188	AlaSerArgAlaThrGlnTyr	-----	204
DB	3646	ATACAAACAAAACAGTTTCGATAAATTTCTACAAAATTTTACTATATGAAGTGAA	3705	
QY	205	TyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLysAsnIle	224	
DB	3706	ATGGACCAATGTAAACAGAGGTACCGAGCTTATCTCAACACAAAAAAGAT	3756	
QY	225	AlaMetAspValLeuAsnTyrValGlyArgAlaAspGlySerThrPheArgThr	244	
DB	3757	-----GCCCAAAAAAAA	3768	
QY	245	AspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGlu	264	
DB	3769	GACATTTGAAGCTAATAATGAG	3807	
QY	265	TyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnGlnMet	284	
DB	3808	GGAAAACTTCAAGCGCTGAGAACGCCAAGCTTGACTTGAAAAAATAATTCANTAGGCTA	3867	

QY	285	LysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGlu	304	
DB	3868	AAGAAGCAGCGCTCATGAAAAATTTGGACGCATCTTAAGAAACAA	3909	
QY	305	AspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGlyGluLeuSerArg	324	
DB	3909	-----	3909	
QY	325	AspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArg	344	
DB	3910	-----CAAGCTGCTTTAACTAACGAACCTCAAC	3936	
QY	345	LysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAla	361	
DB	3937	GAA-----TTAAAGCAATTAAAGACAAATTAGAACCAAGATCTACACTTTGAA	3984	
QY	362	-----SerLysValHisSerGluGlyLeuGluAsp	373	
DB	3985	AACGCTAAAGTTATTGATCTAGATCGAAGCTCAAGCACATGAACCTCAATCAGAAGAC	4044	
QY	374	ValIleSerArgGlnAspHisGluGlnGluThr	384	
DB	4045	GTC-----TCGAGAGATCACGAAAAGGACACTTATCGGCATTAATGGAGGAATTGAA	4098	
QY	385	-----	385	
DB	4099	TCACTCAAAAAGGAGTTACAAAATTTTCAAAACCTGCGAATAGCTCATCTGATGCTTTGAG	4158	
QY	386	LysLeuGluLeuGluIleGluArgCysLysGlu--MetIleLysAlaGlnGlnLeu	404	
DB	4159	AAATTTGAAATTAACATGAGGAAAGAAAAGGACAGAAATTTATTGACGAAAGACGAAGAA	4218	
QY	405	LeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeuLeuArgAspCys	424	
DB	4219	TTTGAAAAAATTAACAGAGACCTTAATTAAGAGTACAAGCTCA	4269	
QY	425	TyrLeuLeuGluGluLysGluArgLysGluGluTrp	437	
DB	4270	GAATACAGTAAAGATATCGAAACATTTGAAAAGGAATGCTAAAGGAATATGAAGATGAA	4329	
QY	438	ThrLeu-----PheLysGluGlnLysLysAsnPheGluArgGluArgSerPheThr	455	
DB	4330	ACGTTAAGGAGAAATTAAGGACAGAGAGATTTAAAGAAAGAAATAGAGTTACCAAGC	4389	
QY	456	GluAlaAlaIleArg-----LeuGlyLeuGluArgLysAlaPheGluGluArgAla	473	
DB	4390	GAGGAAGGATACAAAGATCATTAGTAGCGCTAAAGAGAATTAGAGAAGAAATTTCCGA	4449	
QY	474	SerTrpValLysGlnGln	479	
DB	4450	AAAGAAATTGAAAAGAAAAACGCGGAAGCTTAAACGTTCTTTGTATAACAAAGTAGCGGGAA	4509	
QY	479	-----	479	
DB	4510	GATCGCGAAGAGAATATTGGAACAGTCCATCGAAGGGGAATTTCTGAGAGCGCTCGGCT	4569	
QY	480	-----PheLeuAsnMetThrAsnPheAspHisGlnAsnSer	495	
DB	4570	GTCGCTGGCTTTATTAATCAAAAAACCTTAAACCTCAAGCAACAAATTGAAGATGTGAAG	4629	
QY	496	LeuPheSerAlaPheSerGlySer	510	
DB	4630	AATGACGTTTCTTCAATGACAGCCAGTCAATGGTTTACAAATAAAGAAAAATAATATCGTG	4689	
QY	511	ValHisSerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCys	530	
DB	4690	GATAGTAGCGCTGCTGGTAACCAAGCCCATTTCCAACTTTAGTTTGGTGAAGCATTTCTTC	4749	
QY	531	ThrSerLysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArgGlnThr	550	
DB	4750	TCCTCTAAT---ACATCATCTCTTCAAGT-----TTTCAGAACCCATTCACAGCATCC	4800	


```
Db 661 AAAGAGAGGAATTAGCCTCCAGAGGAATCGTTGAACCATATATGAAGAAAATTTTGCAT 720
QY 176 GluLysAspGluVal
Db 721 GAAAGAGAGGAAGTATTACTAAAGGAACAAGCTCTACTCAATCAGAGGGACGAGAACATC 780
QY 187 IleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsn 206
Db 781 CTT---GAGAGATTAGCTTATGTTACTCTCAGAGAAAAGAGTGGAGAGAGAAAAGAA 837
QY 207 LysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMet 226
Db 838 ATTCTTGAAGCTGAACCGAAGGTTTGTAGAGGAAAATATAAGTTGGAGCTGAANAATG 897
QY 227 AspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLys 246
Db 898 GAGGCAATA----- 906
QY 247 ThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArg 266
Db 907 ---GTTTCAAGAGAGAGAGCTCTGATTTCAGAGGAATCCCTGCTTGAC-----AAAAGG 957
QY 267 GlnLysGlnIleLeuMet
Db 958 GAGAGTGAATTATTAATTTGCAAGAGACAAATCCCAAGTAAAGAACCGGCCGAGATTGAA 1017
QY 279 LysValLeuGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLys 298
Db 1018 AGGTTGAATCAGGAA-----CAAGCTATAGCCTTG----- 1047
QY 299 ProArgGluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAsp 318
Db 1048 ---GAGAGGAGAAAACATGATTTTGAATCTGAGATCGCAAAACAAACAAATGCTTTTGAT 1104
QY 319 SerGly---GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGlu 337
Db 1105 GCAGCAATGGAGGTGACAGAAATGCGCTGCATCAGAGAGAATGT---GCTCTCAGTGAG 1161
QY 338 GlnLeuThrAsnSerIleArgLysGlnTrpArgIle----- 349
Db 1162 CAGGAATCTGTGTTGTGCAAGATCCAGAACTCTTGACCTTCAGCTAGCTGAATTAGCA 1221
QY 350 -----LeuLysSerHisValGluLysLeu 357
Db 1222 AGCAAGAAAACGGCTTGCAGGAGATCAGATGAGTTGAAAGAGAGAGGAGGAAAAGCTC 1281
QY 358 AspAsnGlnAlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArg 377
Db 1282 TTGTTACACAGAGAACCCATACCAATGAATCTTCAAAAAGAAAGGGAAGAAATACAAAGG 1341
QY 378 GlnAspHisGluGlnGluThrGluLysLeuGluIleGluArgCysLysGluMet 397
Db 1342 ATTAATATCGGATTTGAGAGAGAAAAGGCTTCTTTGAGAGAGAGAAA---CGGGAAGCA 1398
QY 398 IleLysAlaGlnGlnLeu-----Leu 405
Db 1399 ATTCAGGCCCAACAAGATCTAGCAATAACACAGCAGATAGAGATGAGTTGCTTACTCTG 1458
QY 406 GlnGlnGlnLeuAlaThrThrCysAspAsp-----AspThrThrSerLeuLeuArg 422
Db 1459 CAGATGAACCTTAAGAGAAATTTGACGCTCAGAGCCCCAAAAGGGAACCTCATGGCT 1518
QY 423 AspCysTyrLeuLeuGlu---GluLysGluArgLeuLysGluGluThrLeuPheLys 441
Db 1519 GATGCAGATAGGCTGCAAGCCGAAAGAGAAAGGTTTGAGATTGAATGGGAGCTGATTGAT 1578
QY 442 GluGlnLysLysAsnPheGluArg-----GluArgArgSerPhe 454
Db 1579 GAAAGAAAGAGGAGCTACAAAAGGAAGCGCATCAGAAATTTGCTGAAGAACGAAGCAATA 1638
QY 455 ThrGlu----- 456
```

```
Db 1639 ACTGAGTATCTGAGAAATGAATCTGATATATCATCAACAGGAGAGGATAATCTCCGTGTT 1698
QY 457 -----AlaAlaIleArgLeuGlyLeuGluArgLysAlaPhe----- 468
Db 1699 CAGTTCAAAGATTAATTCAGAAAACACTCTCTCGTGAACACAAAGAGTTTCATGAGTAAGATG 1758
QY 469 GluGluGluArgAlaSerTrp-----ValLysGlnGlnPheLeuAsnMet---Thr 484
Db 1759 CAGCAAGAACAATGCAAGTTGGCTGAGTAAGATTCAACAAGAAAGGCAAGATCTGAAAGAGA 1818
QY 485 AsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPheSerGlySerSer 504
Db 1819 GACATTGATATCCAGAGGGTGGAA-----TTGCTAAATTCCTGCTAAGGCAAGCCAGATG 1872
QY 505 AspProAspAsnLeuIleValHisArgProArgGlnLysLysLeuHisSerValAla 524
Db 1873 GAAATAGATTCTTATTGT-----AGGGAAGGGAGGAAGAGTTTGAG----- 1914
QY 525 AsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAlaSerProSerThr 544
Db 1915 -----CAGAAAAGGCCAAGGAACTCGAACACATCAATTTCTCAG 1953
QY 545 SerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIleSerValLeuAsn 564
Db 1954 AAGGAGATGATCAACACAAAA-----TTAGAACATGTTGCAGTTGAATTCAGAAA 2004
QY 565 IleThrProGluGluSerLysProSerGluValAlaArgGluSerThrAspGln----- 582
Db 2005 CTT---AAGGATGAGAGAAAAGAGCTACTTTTGGAACTGTGAGAGGAGAGCAAGAGTTG 2061
QY 583 -----LysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGly 599
Db 2062 TCTGAGATAAAGGCCACTATTGAAGCCTTGAATAATCAACGGGAG-----AAG 2109
QY 600 CysSerSerAlaPheArgSerAlaHisGlyAspArgAsp 612
Db 2110 CTGCAAGAGCAAGAAAGAACTATTACATTCAGACCCGAGAA 2148
```

Search completed: June 14, 2006, 02:18:27

Job time : 1992 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 14, 2006, 00:53:30 ; Search time 238 Seconds
(without alignments)
912.379 Million cell updates/sec

Title: US-10-644-084-2

Perfect score: 3165

Sequence: 1 MGDWTVTPVLCTENKLS.....CYSGCSSAFSAHGDRDDLP 615

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 290227 seqs, 117694381 residues

Total number of hits satisfying chosen parameters: 580454

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+p2n.model -DEV=xlp
-Q=/abes/ABSSWEB/spool/US10644084/@CGN_1.1.29@runat_12062006_150434_1613/app_query.fasta.1
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -LOCAL=ALIGN=200 -THR SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=ptco -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -HOST=abss905p
-USER=US10644084 @CGN_1.1.29@runat_12062006_150434_1613 -NCPU=6 -ICPU=3
-NO MAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New.*

1:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US09_NEW_PUB.seq.*
2:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US06_NEW_PUB.seq.*
3:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_NEW_PUB.seq.*
4:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_NEW_PUB.seq.*
5:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
6:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US10_NEW_PUB.seq.*
7:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US11_NEW_PUB.seq.*
8:	/EMC_Celerra_SIDS3/ptodata/2/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	359.5	11.4	1556	US-10-449-902-2430	Sequence 2430, Ap
2	356	11.2	1601	US-10-449-902-27762	Sequence 27762, A
3	349	11.0	1497	US-10-953-349-6159	Sequence 6159, Ap
4	231.5	7.3	5373	US-11-217-529-5516	Sequence 5516, Ap
5	230.5	7.3	5040	US-11-217-529-77055	Sequence 77055, A
6	215.5	6.8	1833	US-11-217-529-80544	Sequence 80544, A
7	209	6.6	2479	US-11-293-697-1966	Sequence 1966, Ap
8	207	6.5	3270	US-10-480-962-24	Sequence 24, Appl
9	206.5	6.5	10211	US-10-505-928-326	Sequence 326, Appl

10	205	6.5	5558	7	US-11-217-529-6026	Sequence 6026, Ap
11	201.5	6.4	5787	7	US-11-217-529-76923	Sequence 76923, A
12	192	6.1	4925	7	US-11-289-102-97	Sequence 97, Appl
13	190.5	6.0	5028	7	US-11-217-529-5947	Sequence 5947, Ap
14	188	5.9	1648	6	US-10-953-349-20044	Sequence 20044, A
15	187	5.9	2773	6	US-11-293-697-2130	Sequence 2130, Ap
16	185	5.8	9588	6	US-10-505-928-101	Sequence 101, App
17	183	5.8	3039	6	US-10-449-902-28253	Sequence 28253, A
18	181	5.7	1684	6	US-10-953-349-20040	Sequence 20040, A
19	181	5.7	2355	6	US-10-449-902-20065	Sequence 20065, A
20	178.5	5.6	1462	6	US-10-953-349-24038	Sequence 24038, A
21	178.5	5.6	2268	7	US-11-217-529-6178	Sequence 6178, Ap
22	178	5.6	2631	7	US-11-217-529-560	Sequence 560, App
23	177	5.6	4263	7	US-11-327-900-5	Sequence 5, Appl
24	176	5.6	3044	7	US-11-293-697-2049	Sequence 2049, Ap
25	175	5.5	2025	7	US-11-217-529-81827	Sequence 81827, A
26	175	5.5	4251	7	US-11-217-529-3111	Sequence 3111, Ap
27	174	5.5	4455	6	US-10-504-973-70	Sequence 70, Appl
28	173.5	5.5	1625	6	US-10-953-349-17058	Sequence 17058, A
29	173.5	5.5	3720	6	US-10-449-902-17102	Sequence 17102, A
30	173	5.5	1798	7	US-11-293-697-1637	Sequence 1637, Ap
31	172.5	5.5	1647	7	US-11-217-529-3824	Sequence 3824, Ap
32	171.5	5.4	2466	7	US-11-217-529-78674	Sequence 78674, A
33	170	5.4	1257	7	US-11-217-529-80717	Sequence 80717, A
34	169	5.3	1380	7	US-11-293-697-2304	Sequence 2304, Ap
35	169	5.3	3735	7	US-11-217-529-2189	Sequence 2189, Ap
36	168.5	5.3	2607	7	US-11-217-529-4263	Sequence 4263, Ap
37	168	5.3	2930	7	US-11-293-697-643	Sequence 643, App
38	167.5	5.3	1928	7	US-11-293-697-1901	Sequence 1901, Ap
39	167	5.3	2477	6	US-10-953-349-11093	Sequence 11093, A
40	165	5.2	1203	7	US-11-217-529-79694	Sequence 79694, A
41	165	5.2	2520	7	US-11-217-529-1961	Sequence 1961, Ap
42	164.5	5.2	2796	7	US-11-217-529-81089	Sequence 81089, A
43	164	5.2	2072	6	US-10-953-349-11127	Sequence 11127, A
44	163.5	5.2	1717	6	US-10-449-902-2456	Sequence 2456, Ap
45	163.5	5.2	3374	6	US-10-449-902-23271	Sequence 23271, A

ALIGNMENTS

RESULT 1
US-10-449-902-2430
; Sequence 2430, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-Oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2430
; LENGTH: 1556
; TYPE: DNA
; ORGANISM: Oryza sativa
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK060264
; DATABASE ENTRY DATE: 2001-12-06
US-10-449-902-2430
Alignment Scores:
Pred. No.: 1.13e-23 Length: 1556
Score: 359.50 Matches: 105
Percent Similarity: 48.7% Conservative: 97
Best Local Similarity: 25.3% Mismatches: 163


```
QY 115 LeuAlaGlnGluSerValGluThrGlnAsnLeuLeuGlySerAspMetAspHisLeu 134
Db 333 GAGTTCGGAGATCTACCAATGATCTCGCTCAGCTGATGATATTTCTAGATTA 392
QY 135 GlnSerCysTyrAlaLeuLeuGlySerGluThrSerArgArgGluMetIleGly 154
Db 393 GAAGCCAAAGATTGAAGAATGGATGCTCAATTAGCTGCAAAAGATCGCGAGTGGCCACA 452
QY 155 LeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLys 174
Db 453 TTGACTAGAACGGAGCCCAAAACACTGCAACTTTTGAAGGCTCAGATTGACAAAGTGCAC 512
QY 175 AsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArg-----AlaThr 192
Db 513 CAGGAACGGATGAGTTCCAAAATATG-----GTTATTGGCAATCAGCAAGTACGTACC 566
QY 193 GlnTyrAsnHisAspValLysArgLysGluArgGlyThrAsnLysLeuLysGluArgLeu 212
Db 567 CAGCAAAATTCATGAATGAAGAAAGAAAGAAAGATACATCAATTCGAGGAGAGTTA 626
QY 213 HisGlnLeuValMetAsnLysLysAspLysAsn-----IleAlaMetAspValLeuAsn 230
Db 627 AACCCAGGTATTATGAGAGAAAGAAAGGAATCATCAGTTTCAGGAATGGAATATGAAC 686
QY 231 TyrValGlyArgAlaAspGlyLysArgGlySerTyrArgThrAspLysThrGluAlaArg 250
Db 687 TTGTTGCAGAAAGAGGACGGCAGCGTGGAACTTTGGAGTGGAAAAAG-----734
QY 251 AsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIle 270
Db 735 NATGCAATGATTATTACAAATGATTGTTGATGCTATGAGTAAAGAAAGCAAGATTG 794
QY 271 LeuMetGluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetLysLeu 290
Db 795 ATGCAAGAAATGCTGATTGCGAGCACTATTGCGTTCCGATGCAGATGATATGCGCAG 854
QY 291 LeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGlyThrGlyThrVal 310
Db 855 TTCCTCAAT--GCTCCAAATGGAGTTCAACAGCCAGCTGTTGTGGGCAATCGACACAG 911
QY 311 AlaIleSerAspIleGluAspSerGlyGluLeuLeuSerArgAspSerValTyrGlyLeu 330
Db 912 GAGGCAGGATCTCCTCAGTCTCCACTTGT-----GGCAACAGGATGCTTTGATTG 965
QY 331 SerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTyrArgIleLeu 350
Db 966 CCCTTTACATGGCTAGAGACCAAGATTGAAGAGAGTTTTCGCGCACTAAATATGCTTCCATA 1025
QY 351 LysSerHisValGluLysLeuAspAsn-----GlnAlaSerLysValHisSerGluGly 368
Db 1026 AAGGCCCGAATGACACAACTTCAAGATGCCAAAAGGGTGCAGAAAGTGCACATCAGAAGCT 1085
QY 369 LeuAsnGluGluAspValIleSerArgGlnAspHisGluGlnGlnLeuGluLysLeuGlu 388
Db 1086 ACC-----GACCGGGAAGTTGAA---CTTGAA 1109
QY 389 LeuGluIleGluArgCysLysGluMetIleLysAlaGlnGlnLeuGlnGln 407
Db 1110 GCTCAACTAGTTGAAGCTAGAGCATTATTCAGGAGCAGGCATCCTCATGTGCCAAG 1166

RESULT 3
US-10-953-349-6159
; Sequence 6159, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nickolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
```

```
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6159
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-953-349-6159

Alignment Scores:
Pred. No.: 9,53e-23 Length: 1497
Score: 349.00 Matches: 110
Percent Similarity: 50.0% Conservative: 102
Best Local Similarity: 25.9% Mismatches: 148
Query Match: 11.0% Indels: 65
DB: 6 Gaps: 13
```

US-10-644-084-2 (1-615) x US-10-953-349-6159 (1-1497)

```
QY 2 GlyAspThrMetThrValThrAspProValLeuLeuCysThr-GluAsnLysAsnLeuSerG1 21
Db 90 GCGCGCTACACACCGGTTATCGATGTCATTGATGCTGACCGAG-----132
QY 21 nTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCysSe 41
Db 133 -----TTCGATTTTCAGATTTCAGCCACACAGTTTGGTTTCTCACAGTCA-----174
QY 41 rSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGluAs 61
Db 175 -----CCGATCGGGAGCGT-----AATTTCCGGATGTTGTATAA 209
QY 61 nIleGlnGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSerLe 81
Db 210 TCTGAGAACTGTTAAGTATTGTAATCAGAGCTTAGTCACCTCTGGATTCTCTCT-- 267
QY 81 uTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeuAs 101
Db 268 -----TCGCTTGACCTCTTCGCTACCGATCCTGTTTCGATTGTAGAATTCGCAA 317
QY 101 nCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuAlaGlnGluSerValG1 121
Db 318 TTGCGTATACGCTTTGATTCAACAGAGCAACAGATGTGGAATTTAGAGAAATCTGCTAA 377
QY 121 uThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnGlnSerCysTyrAlaLysLe 141
Db 378 TGATCAGAGACAAACGACTTTTATCCGATATGCGAGACTCGAAGCAAAAGTTGAAAGGCT 437
QY 141 uLysGluGlnLeuGluThrSerArgArgGlu-MetIleGlyLeuGlnGluArgAspArg 161
Db 438 TGAGACTCAATTGCAAGCTAAAGAAAGGGAACCTTGGTTCCGTGACAAAGAACGACAAA 497
QY 161 nLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGlnLysAspGluValG 181
Db 498 AAATACAGCG-GCTTTAAAGACGCAAGATGAGAAGCTTCAAAAAGAGAGAGATGAGTTTC 556
QY 181 nLysLeuGlnAsnIleAlaSerArg-----AlaThrGlnTyrAsnHisAspValL 199
Db 557 AACGAATG-----GTGATTGCTAACCCAGCAAGTCAAGACCAACAAATTCATGAAACAA 610
QY 199 ysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMet---A 218
Db 611 AAAAGAAAGAAAGAGAGTACATAAAGTTACAGGAGAGGTTAAACCAAGTTGTTGATGGAGA 670
QY 218 nLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyL 238
Db 671 AAAAGAAAGAAAGAGAGTACAGGCATGGAGATTATGAATCTACTCCCAAGAAAGAGGAGAC 730
QY 238 ysArgGlySerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysI 258
Db 731 AGCGTGGAACTGGAGTGGGAGAAAGAACTGACTCT-----GATTTCACAAA 778
QY 258 leLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeu 278
Db 779 AAATTTGATGTCATATGCGGCAAAATCAAGAACTGATGCGCAGAAACACTGATCTGA 838
```



```

QY 308 GlyThrVal-----AlalIleSerAspIleGluAspAspSerGlyGluLeuSerArgAsp 325
Db 4231 AGACTAATTGATACAAAGTAGAGTGACCTGGAGAAAGCCACT----- 4272
QY 326 SerValTrpGlyLeuSerCysAspThrValArgGluLeuThrAsnSerIleArgLys 345
Db 4273 -----TTGAGCCATGATGAATTTTAGAGAAAGGCGGATGAGATTATAAA 4320
QY 346 GlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysVal--- 364
Db 4321 ATGAAAGACGAAATCTCTCATCAAAAGAAAGATAAATGAAGCAAAACACGAAACTACTA 4380
QY 365 -----HisSerGluGlyLeuAsn-----GluGluAspValIleSerArg 377
Db 4381 TCAACGGAAACACCATGAATAATGGGTAAATAGTTTGAAGGAACAACCTTGAGGCTATT 4440
QY 378 GlnAspHisGluGlnGluThrGluLeuGluGluLeuGluArgCysLysGluMet 397
Db 4441 CAAGAATCTAATCCGAGGTTGAAGAGAAATTAAGAGACATGAGGAAAAAATCTGCTCAC 4500
QY 398 IleLysAlaGlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspThr 417
Db 4501 CTCAAATCAGATTTAAAAAATCAAAAGAAACTGCGAAAGACTTAAAAAGTGACATTGAA 4560
QY 418 ThrSerLeuLeuArgAspCysTrpLeuLeuGluGluGluArgLeuLysGluGluTrp 437
Db 4561 AACATGTAAAGAAAGAAATAGAACTTTTAGTAAGAGTGGGAAGGAATCTGATGAGAAGTTT 4620
QY 438 ThrLeuPheLysGlu-----GlnLysLysAsn 446
Db 4621 AACAGTATAAGAAATAATAGTCAGGTAGACATTAAAGAGTTGCAAGATGAAAGACTGAT 4680
QY 447 PheGluArgGlu-----ArgArgSerPheThrGluAlaAlaIleArgLeuGly 462
Db 4681 TTGGAATCTCAATCGTAGAATTCAGAAAAAATCGAAGAGCTACAACTGAAA----- 4734
QY 463 LeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPheLeuAsn 482
Db 4735 CTGAAAGACGCAAGCAGAAATCTAACAGTAAGATAGACACCATACAGCAAGAGCTAACTAAT 4794
QY 483 MetThrAsnPheAspHisGlnAsn---SerGluAsnValLysLeuPheSerAlaPheSer 501
Db 4795 GCTTATGAGAAAATAAGAGTTAATGAAGAGAAATGTATTATTA----- 4839
QY 502 GlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeuHis 521
Db 4840 ---AACTCCAAATTGAAAAAATTTAGAAATGCAATTTAAGAGATAATCAAGACGAGATACAG 4896
QY 522 SerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla--- 540
Db 4897 AGC-----ACTAAGGAAAAGAACGCAATCATTTGGATTCTTAC 4932
QY 541 -----SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGlu 556
Db 4933 TTGAAGATCTGAAGCAGAGTTGACTGACGCTCAACAAAAGCGCTAAAGAAATCGCAAGAA 4992
QY 557 HisSerSerIleSerValLeuAsnIleThrProGluLysSerLysProSerGluValAla 576
Db 4993 AAGAGTAAACCGGAAATTTGAACAACCTTCAAGCAGAGAGCGCTCTCGATCTCATGAATGCC 5052
QY 577 ArgGlu-----SerThrAspGlnLysTrp 584
Db 5053 AAACAACACTGAAATTAATAATACAGTAACCTTGAACCGAAGACAAACTTGG 5103

```

RESULT 5

```

US-11-217-529-77055
; Sequence 77055, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA

```

```

; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 77055
; LENGTH: 5040
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-77055

Alignment Scores:
Pred. No.: 2,75e-11 Length: 5040
Score: 230.50 Matches: 147
Percent Similarity: 36.0% Conservative: 120
Best Local Similarity: 19.8% Mismatches: 230
Query Match: 7.3% Indels: 245
DB: 29 Gaps: 29

US-10-644-084-2 (1-615) x US-11-217-529-77055 (1-5040)

QY 15 GluAsnLysAsnLeuSerGlnTyrThr----- 23
Db 3028 GAAATTAAGATCTATCTGAATTTGGTTATCCGCTTGAAGAGTGGCTGATGCCAA 3087
QY 24 SerGluThrLysMetSerProSerSerLeuTyrSerGlnValLeuValPheCysThrGlyGlu 43
Db 3088 GCAGAAATTAACAAAAACCAATCCCTCATTTGTATAGCACTCAGGATTTA----- 3135
QY 44 ProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu----- 60
Db 3136 ---CTCGATAAACATGAAAGGAAATGATGGAGGAAAGGCTCATTTATGAAAGAGAATTG 3192
QY 61 -----AsnIleGluGlnSer----- 68
Db 3193 ATATCTAATATTGAACAGACGGAATCCTTCGCTGTGAAAAATTCGGTATTGATGAAAG 3252
QY 69 LeuAspGlnGluLeuThrThrPheGly-----Phe 78
Db 3253 GTTGATGACACAGACGAAATTAATGCGATAAAGATCATTTGAAGCTAGTTCTTTGTTT 3312
QY 79 ProSerLeuTyrGluLysSerLysSerLysGluAlaLysArgGluLeuAsnIleValAla 98
Db 3313 TCAAACTTACGTCATGAAAGAAATCTTTTAGAGACGAAAG----- 3351
QY 99 ValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuAlaGlnGlu 118
Db 3352 TTAAACCACTGCAAGAGAGAAATTCGATTCGTTAAACAAAAGACGATAGTTTAGAAAAA 3411
QY 119 SerVal---GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCys 137
Db 3412 ACGATAAATGACCTCAACGAAACCCAGACTTTTATCGAGAGGAATAC-----CAATGC 3465
QY 138 TyrAlaLysLeuLysGluGlnLeuThrSerArgArgGluMetIleGly----- 154
Db 3466 TCTGCGATTATTATTGACGAATTCAGAGTATATCAGAAAGAGTTACTCAGGTTAACATT 3525
QY 155 -----LeuGlnGluArgAspArgGlnGlnCysLysAsnArg 167
Db 3526 TTAAAGAGATAATGCCATCTTACAAAGAGCTCTCAAAAACGTAACCGAAAAAAGACAG 3585
QY 168 SerLeuHisGlnLeuLysAsnGlnLysAspGluValGlnLysLeuGlnAsnIleIle 187
Db 3586 GAAATTTATAACAATTAATGATAGGCAAGAGAAATTTCTCGATTACAGCGTGATCTC 3645
QY 188 AlaSerArgAlaThrGlnTyr-----AsnHisAspValLysArgLysGluArgGlu 204
Db -----AsnHisAspValLysArgLysGluArgGlu 204

```

[illegible]

Qy	480	-----PheLeuAenMetThrAsnPhesHisGlnAenSer---GluAen	493
Db	4564	TCGCCTGTCACTGGCTTTTATAAAATCAA AAAAACCTTAACCTCAAGAACAAATTGAAGAAT	4623
Qy	494	VallYsLeuPheSerAlaPheSerGlySer-----SerAspProAspAen	508
Db	4624	GTGAAGAATGACGTTTCCTTCAATGACAGCCAGTCAGTCAATGGTTACAATAAAGAAAAATAAT	4683
Qy	509	LeulleValHisSerArgProArgGlnLysLysLeuHisSerValAlaAenGlyValPro	528
Db	4684	ATCGTGATAGTAGCGCTGCTGGTAAACAAGGCCAATTCACACTTTAGTTTTGGTAAAGCCA	4743
Qy	529	AlaCysThrSerLysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArg	548
Db	4744	TTCTTCCTCCTTAAT--ACATCATCTCTTCAAAGT-----TTTCAGAACCCATTCA	4794
Qy	549	GlnThrHisSerCysValSerGluHHisSerIleSerValLeuAenIleThrProGlu	568
Db	4795	GCATCCCAATCTAATATATCAATACGAATGCCCCCTAGCTACACTCAATATTTCACCTGAA	4854
Qy	569	Glu-----SerLysProSerGluValAlaAeGluSerThr	580
Db	4855	GTTCAGCTCAAAAGCGCGATTAACTTTTCCAACGTAAACCAGTCTAACTTAACATAGCAC	4914
Qy	581	AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCys	600
Db	4915	CATCGAGCAAAAATAACAGAAATGGAAAGTACATCTAAACGGCCAATTGAGAGCGGTACA	4974
Qy	601	SerSer 602	
Db	4975	TCITCT 4980	
 RESULT 6			
US-11-217-529-80544			
; Sequence 80544, Application US/11217529			
; Publication No. US20060099612A1			
; GENERAL INFORMATION:			
; APPLICANT: SUNTORY LIMITED			
; APPLICANT: NAKAO, YOSHIHIRO			
; APPLICANT: NAKAMURA, NORIHISA			
; APPLICANT: KODAMA, YUKIKO			
; APPLICANT: FUJIMURA, TOMOKO			
; APPLICANT: ASHKARI, TOSHIOKO			
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS			
; FILE REFERENCE: S-38-285			
; CURRENT APPLICATION NUMBER: US/11/217,529			
; PRIOR FILING DATE: 2005-09-02			
; PRIOR APPLICATION NUMBER: US 10/932,182			
; PRIOR FILING DATE: 2004-09-02			
; NUMBER OF SEQ ID NOS: 197023			
; SOFTWARE: PatentIn version 3.3			
; SEQ ID NO 80544			
; LENGTH: 1833			
; TYPE: DNA			
; ORGANISM: Saccharomyces pastorianus			
US-11-217-529-80544			

RESULT 6

```

US-11-217-529-80544
; Sequence 80544, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 80544
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-80544

```

Alignment Scores:		
Pred. No.:	1.53e-10	Length: 1833
Score:	215.50	Matches: 123
Percent Similarity:	39.1%	Conservative: 109
Best Local Similarity:	20.7%	Mismatches: 217
Query Match:	6.8%	Indels: 145
DB:	7	Gaps: 24

US-10-644-084-2 (1-615) x US-11-217-529-80544 (1-1833)

Qy	62	IleGluGlnSerIleSerTyrLeuAspGlnGluLeuThr-Thr-PheGlyPheProSerLeu
		::: ::
Db	4	TTGAGACAAAAATTCCTTTTGATGGATGTTGAATTA-----
Qy	82	TyrGluGluSerIlysserLys-----GluAlaLysArgGluLeuAsn

```

Db      40  ---CAAGATGCTCGTACTAGCTAGATAATTCAGAGTTGAAAAGGAAATCAATCTTCC 96
Qy      96  ILeValala-----ValLeuAsnCysMetAsnGluLeuLeuValLeuGlnArg 111
Db      97  ATTATTCAACAGCATGACGACATATTATGGAGAAATTAATCAATTAATCTATTAAAGAA 156
Qy     112  LysAsnLeuLeuAlaGlnGluSerValGluThrGlnAsnLeuLys-----LeuGly 128
Db     157  AGTAAATACATACATTGCGGAATGAGCTGGAAAAACAACAATAACAGAGAAAGGAACATGCAA 216
Qy     129  SerAspMetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGluThrSer 148
Db     217  TCTGAATTAGATAAATTGAAGCAAAATGTTGCGCTATCGAGTCCGAATTCAGCTGCTTG 276
Qy     149  ArgArgGluMetIleGlyLeuGlnGluArgAspArgLeuGlnCysLysAsnArgSer 168
Db     277  AAATATTCTATG-----CAAGAAAAAGCAAGAGAGCTCAATATTAGCTAAAGAAGAG 327
Qy     169  LeuHisGlnLeuLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAla 188
Db     328  GTTCATCGTTGGAAAAAGCGCTCACAGACATATTGGAGAACATGAACAATTTGAGCTCA 387
Qy     189  SerArgAlaThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeu 208
Db     388  AGC-----GATTATGAGAGCTAGAAAGCGAGATAGAAAATTTG 426
Qy     209  LysGluArgLeuHisGlnLeuValMetAsnLysLysAspLysAsnIle---AlaMetAsp 227
Db     427  AAGGAGGAACCTAGAA-----AATAAGGAGCGTCAAGGAGCGGAAGCCGAGGAA 474
Qy     228  ValLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThr 247
Db     475  AAATTTACAGGCTGAGAGACAAAGCGCAGGAGAGA-----TTAAAAACATCAAAAATC 528
Qy     248  GluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGln 267
Db     529  TCACAGGACTCATGCTGACACAAAGTAAATAGTCTTAAGGGATGCAAGAAGACGTGTGAA 588
Qy     268  LysGlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGlu 287
Db     589  AATTCCTTGAGTGAGCAAAACGCGAGAAATCGAAGAG---TTACAAAATCAAAAAGTA--- 642
Qy     288  MetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGlyThr 307
Db     643  -----GCACAAGGTACAAC 657
Qy     308  GlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeuSerArgAspSerVal 327
Db     658  CAGTTAGAGCAATAGAAAATTAACAAGAAGCGCAGAAAAGGCTTCAAGAGAGCTTCAA 717
Qy     328  TrpGlyLeuSerCysAspThrValArg---GluGlnLeuThrAsnSerIleArgLysGln 346
Db     718  GCCAAGTTAGAGAAAGTACGACTCTTACGAATCTACGATAAACGGCTTAAATGAAGAA 777
Qy     347  TrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSer 366
Db     778  ATTACAACATTTAAAGAAGAAATTTGAAAAAACAAGGCAAAATCCAGCAACAGTTACAAGCT 837
Qy     367  GluGlyLeuAsnGluGluAspValIleSerArgGlnAspHisGluGlnGluThrGluLys 386
Db     838  ACATCTGCAAAATGACAAAATGACTTATCT-----AACATAGTTGAATCTATGAAA 888
Qy     387  LeuGluLeuGluIleGluArgCysLysGluMetIleLysAlaGlnGlnLeuLeuGln 406
Db     889  AAGTCTTTTGAAGAGATAAATAATCAAA---TTTCATCAAAAGAAAAAACCCCAAGAGTTAAT 945
Qy     407  GlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeuLeuArgAspCysTyrLeu 426
Db     946  GAAAA-----ATA 954
Qy     427  LeuGluGluLysGluArgLeuLysGluGluTyrThrLeu---PhelysGluGlnLysLys 445

```

```

Db      955  CTTGAGGCCCAAGAAAGGCTTAATCAACCTTCCAATATCAATATGAGGAGATTAATAAAA 1014
Qy     446  AsnPheGluArgGluArg----- 451
Db     1015  AAATGGGAATCTGAGCAACGAGGAAGTATCTCAAAAAGATTTCGCAAGCTGAGGAAGCC 1074
Qy     452  -----ArgSerPheThrGluAlaIleAileArgLeuGlyLeuGluArgLys 466
Db     1075  CTCAAAAACGAATCAGATTTACCCACTGAGGAGAAAATCAATAAGATAATTCGAAACGAAAG 1134
Qy     467  AlaPheGluGluGluArgAlaSerTrpValLysGlnPheLeuAsnMetThrAsnPhe 486
Db     1135  AAGGAGGAATTTGGAAAA-----GAGTTT 1158
Qy     487  AspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPheSerGlySerSerAsp--- 505
Db     1159  GAAGAAAAGGTTGAGGAGAGAATAAAATCAATGGAACAA---TCTGGAGAAATAGACGTG 1215
Qy     506  -----ProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeuHisSerVal 523
Db     1216  GTGCTTCGAAAAACAGCTAGAAAGCTAAGGTTCAAGAGAGAAACAAAGGAATTTGGAACGAG 1275
Qy     524  AlaAsnGlyValProAlaCysThrSerLysLeu-----ThrLysSerLeuProAla 540
Db     1276  TATAC-----AAAAATTACAAGAAGAACTCAAGATGTCCACAC 1317
Qy     541  SerProSerThrSer----- 545
Db     1318  TCCCCTCATATCTCAGGTCATGAGAGGAGCAAGATTACGGCGCAGAAAATCGAAAGCAGTTG 1377
Qy     546  -----AppPheArgGlnThrHisSerCysValSerGluHisSer----- 558
Db     1378  AGGGAGGAGTTCAACAATGAACCTGCAAGCCATAAAGAGAAAATCTCTTCGACGAGGAAG 1437
Qy     559  SerIleSerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGlu 578
Db     1438  CAACAAAGCAATGATGAAAACCTACCCCTTTTGGAAAGAAAACCTTGCCAAAGATGGATCTCAA 1497
Qy     579  SerThrAspGlnLysTrpSerValGlnSerArgProSerSer 592
Db     1498  TTGTGAGAAACAAACAAAGTCCGAGACTCTCCGAAATCT 1539

RESULT 7
US-11-293-697-1966
; Sequence 1966, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1966
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1966

Alignment Scores:
Pred. No.:          9,06e-10      Length:      2479
Score:              209.00        Matches:    108
Percent Similarity: 42.4%         Conservative: 113
Best Local Similarity: 20.7%      Mismatches:  192
Query Match:        6.6%          Indels:     109
DB:                  7            Gaps:       20

US-10-644-084-2 (1-615) x US-11-293-697-1966 (1-2479)

Qy     12  LeuCysThrGluAsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSerProSer 31

```



```
; SOFTWARE: PERL Program
; SEQ ID NO 24
; LENGTH: 3270
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 3418524CB1
US-10-480-962-24

Alignment Scores:
Pred. No.: 2,02e-09 Length: 3270
Score: 207.00 Matches: 138
Percent Similarity: 37.1% Conservative: 117
Best Local Similarity: 20.1% Mismatches: 254
Query Match: 6.5% Indels: 179
DB: 6 Gaps: 28

US-10-644-084-2 (1-615) x US-10-480-962-24 (1-3270)

QY 30 ProSerSerLeuTyrSerGlnValLeuCysSerSerValProLeuSerIlys----- 47
|||||
D 741 CCCTCCTGTGAGTATGCG-----ATCCACTGTACCTCACTCCCTTCGTCAGCGCAGAGA 794
|||||
QY 48 ---AsnValHisGly----- 51
|||||
D 795 TAACAACATCATGGATCTGCAGACACAGCTGAAGGAAGTATTGAAGAAATGATCTCTT 854
|||||
QY 52 ValPheGlyValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyrLeuAspGln 71
|||||
D 855 GCGGAAGGATGTGGAAGTAAGAGAGAGCAAAATTGAG-----TTCITCAATGAA 902
|||||
QY 72 -GluLeuThrThrPheGlyPheProSerLeuTyrGluGluSerLysSerIlysGluAlaLys 91
|||||
D 903 TAGCATCAAGACCTTCTGGAGCCACAGAGCTG-----AAGAAGGAACAGCCCT 950
|||||
QY 91 sarGluLeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuValLeuGlnAr 111
|||||
D 951 GAGAAAGATGAAGCTTCCAAATACCAATTTGGAAGGAACAGTACAGAGATTGTACAGGA 1010
|||||
QY 111 gLysAsnLeuAlaGlnGluSerValGluThr-----GlnAsnLeuLysLeuGlySe 129
|||||
D 1011 GGAACACCGACCATGCAGATGACATCCAGGCTCTCCAGGATGAATTCGGATCCAGAG 1070
|||||
QY 129 rAspMetAspHisLeu-----GlnSerCysTyrAl 139
|||||
D 1071 GGACCTGAATCAGCTGTTCAGCAGGATAGTACAGCAGGACTGGCGAACCTTGTGTAGC 1130
|||||
QY 139 aLysLeuLysGlu-----GlnLeuGluThrSerArgArgG 151
|||||
D 1131 AGAGCTGCAGCAGAGGAAACTTTCAGAGGCTTCATGCTGAGCATGAGCGGCGGCCAAAGA 1190
|||||
QY 151 uMetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisG 171
|||||
D 1191 GCTGTTTCTTTCGAAGACATGGAGGAATGGAGCTGCGTATTGAGACTCAAAAGCA 1250
|||||
QY 171 nLeuLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleIleAlaSer----- 189
|||||
D 1251 GACCCTAAATGCTCGGGATGAATCCATTAAAGAGCTTCTGGAATGTTTCGACAGCAAGG 1310
|||||
QY 190 -----ArgAlaThrGlnTyrAsnHisAsp----- 197
|||||
D 1311 ACTTTCGCAAGGCTACCGAGGAAGACCATGAGAGAAACAGACGACTGGCAGAGGCAGA 1370
|||||
QY 198 -----ValLysArgLysGluArgGluTyrAsnLysLe 208
|||||
D 1371 GATGCAGTTTCATCCTAGAAAGCCCTTTGGAGCAGAGGAAGAAAGAAACAGATGATGTT 1430
|||||
QY 208 uLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspVa 228
|||||
D 1431 GAGAGAGGAGATGATCGAGGTTTGAGATGCTCCTGATTCTGCCCAAAACAAAGCTCT 1490
|||||
QY 228 lLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrG 248
|||||
```

Db 2520 AGCA-----TTGGAAGCCAGAGCCAGTCCAGA 2546
 QY 543 rThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIleSerValle 563
 Db 2547 GATGAGTACCGAATACAG-----CACTTGAGAGAGAGATCAC 2585
 QY 563 uAsnIleThrProGluGluSerIysProSer-----GluValAl 576
 Db 2586 CAGGTACAAAGATGAATCTAGCAAGGCCAGGAGGAGTGTGATCGACTCTTAGAATCTT 2645
 QY 576 aArgGluSerThrAspGlnIys 583
 Db 2646 GAAGGAGGTGGAAAATGAGAAG 2667

RESULT 9

US-10-505-928-326
 ; Sequence 326, Application US/10505928
 ; Publication No. US20060088532A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ludwig Institute for Cancer Research et al.
 ; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
 ; FILE REFERENCE: 28967/39178
 ; CURRENT APPLICATION NUMBER: US/10/505,928
 ; CURRENT FILING DATE: 2004-08-27
 ; PRIOR APPLICATION NUMBER: US 60/363,019
 ; PRIOR FILING DATE: 2002-03-07
 ; NUMBER OF SEQ ID NOS: 866
 ; SOFTWARE: PatentIn 3.2
 ; SEQ ID NO 326
 ; LENGTH: 10211
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-505-928-326

Alignment Scores:

Pred. No.: 1,09e-08 Length: 10211
 Score: 206.50 Matches: 156
 Percent Similarity: 35.8% Conservative: 107
 Best Local Similarity: 21.2% Mismatches: 241
 Query Match: 6.5% Indels: 231
 DB: 6 Gaps: 33

US-10-644-084-2 (1-615) x US-10-505-928-326 (1-10211)

QY 17 LysAsnLeuSerGlnTyrThr----- 23
 Db 1657 AGAATACCTCTCAGAAACCATGTTAAGAGATCTTCAAGAAAAATAAATCAGCAAGNA 1716
 QY 24 -----SerGluThrLysMetSerProSerSerLysLeuTyrSerGlnVal 38
 Db 1717 AACTCTTGACTTTAGAAAACTGAAGCTTGCTGGCTGATCTGGAAAAAGCAGCGAGAT 1776
 QY 39 LeuCysSerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThr 58
 Db 1777 ---TGTTCTCAAGACCTTTGAAGAAAGAGAACATCACTT----- 1815
 QY 59 GlyGluAsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPhe 78
 Db 1816 ---GAACAACTTAATGATTAAGTTAAGCAAGACAGAGAAAGAGTCCAAAGCCCTTG---CTG 1869
 QY 79 ProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAla 98
 Db 1870 AGTGCTTTAGAGTTAAAAAAGAAAGAAATATGAATTGAAGAGAGAGAAAACTCTGTTTTCT 1929
 QY 99 ValLeuAsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlu 118
 Db 1930 TGTGGAAAAAGTGAAAACGAA-----AAACTTTTAACACTCAGATG 1968
 QY 119 SerValGluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyr 138
 Db 1969 GAATCAGAAAGGAAACCTTGAG-----AGTAAATTAATCACTTGGAAACTTGT--- 2019
 QY 139 AlaLysLeuLysGluGlnLeuGluThrSerArg-----ArgGluMetIleGlyLeuGln 156

Db 2020 -----CTGAAGACACAGCAAAATAAAAAAGTCATGAATACAACGAGAGAGTAAGAAACGCTG 2073
 QY 157 GluArgAspArgGln---LeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu----- 173
 Db 2074 GAGATGGACAGAAAAACCTAAGTGTCCAGATCAGAAAACCTTCACAACTGTTAGACAGT 2133
 QY 174 LysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArgAlaThrGln 193
 Db 2134 AAGTCAGTGGAGGTAGAGACCAGAAAACCTAGCTTATATGGAGCTACAGCAGAAAGCTGAG 2193
 QY 194 TyrAsnHisAspValLysArgLysGlu----- 202
 Db 2194 TTCTCAGATCAGAAAACATCAGAGAAGAAATAGAAAATATGTGTTTGAAGACTTCTCAGCTT 2253
 QY 202 ----- 202
 Db 2254 ACTGGGCAAGTTGAAGATCTAGAACACAAAGCTTCAGTTACTGTCAAAATGAATAATGAGAC 2313
 QY 203 -----ArgGluTyrAsnLysLeuLysGluArgLeu--- 212
 Db 2314 AAAGACCGGTGTATTACCAAGACTTGCATGCCGAATATGAGAGCCTCAGGGATCTGCTAAAA 2373
 QY 213 -----HisGlnLeuValMetAsnLysLysAsp 221
 Db 2374 TCCAAGATGCTTCTCTGCGTGCACAAATGAAGATCATCAGAGAAGTCTTTTGGCTTTTGTAT 2433
 QY 222 LysAsnIleAlaMet-----AspValLeuAsnTyrValGlyArgAlaAspGlyLysArg 239
 Db 2434 CAGCAGCCTGCATGCATCATCTCTTTGCCAATATAATT-----GGAGAACA 2481
 QY 240 GlySerTyrArgThrAspLysThrGluAlaArgAsnGluAsp----- 253
 Db 2482 GGAAGCATGCCTTCAGAGAGGAGTGAATGTCGTTTAGAAGCAGACCAAGATCCGAAAAAT 2541
 QY 254 -----GluMetTyrLysIle 258
 Db 2542 TCTGCCATCTACAAAAATAGAGTTGATTCACTTGAATTTTCATTAGAGTCTCAAAAAACAG 2601
 QY 259 LeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLys 278
 Db 2602 ATGAACCTCAGACCTGCAAAACGAGTGTGAGAGTGTGGTGCAAATCAAAAGGAGNAATAGAA 2661
 QY 279 LysValLeu-----GlnGlnMetLysLysGluMetIleSerLeuLeuSerProGln 295
 Db 2662 GAAAACTCTAGAAAGCAGAACAGATGCATCAAAAGTTTGTGGCTGAAACAAAGTCAGCGC 2721
 QY 296 LysLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAla-----IleSer 313
 Db 2722 ATTAGTAGTTACAGGAGACACTTCTGCTCACCAGAAATGTTGTTGCTGAAACCTTAAGT 2781
 QY 314 AspIleGluAspAspSerGlyLysLeuSer-----ArgAspSerValTyr----- 328
 Db 2782 GCCCTTGAGAAACAAGGAAAAAGAGCTGCACCTTTAAATGATAGGTAGAAACTCAGCAG 2841
 QY 329 -----GlyLeuSerCysAspThrValArgGlu--- 337
 Db 2842 GCAGAGATTCAGAATTAATAAAGACAGCAACCATCTACTTGAAGACTCTCTAAAGGAGCTA 2901
 QY 338 GlnLeu-----ThrAsnSerIleArgLysGln-----TrpArgIle 349
 Db 2902 CAACCTTTATCCGAAACCTTAAGCTTGGAGAGAAAGAAATAGAGTTCCATCATTTCTCTA 2961
 QY 350 LeuLysSerHisValGluLysLeu-----AspAsnGlnAlaSerLysValHisSerGlu 367
 Db 2962 AATAAAGGGAATTTGAAGAGCTGACCCAGAGAGATGGGACTCTTTAAGGAAATTAATGCA 3021
 QY 368 GlyLeuAsnGluGluAspVal----- 374
 Db 3022 TCCTTAATCAAGAGAGACATGAATTAATCCAGAAAAAGTGAGAGTTTGTCAAACTATATA 3081
 QY 375 -----IleSerArgGlnAspHisGluGlnGluThrGluLysLeuGlu 388

3082 GATGAAAGGAGAGAAAGCATTTTCAGAGTTATCTGATCAGTACAGCAAGCAAGAAACTTATT 3141
389 LeuGluIleGluArgCysLysGluMetIleLysAlaGlnGlnLeuLeuGlnGln--- 407
3142 TTA---CTACAAAGATGTGAAGAAACCGGAAATGCGATATGAGGATCTTAGTCAAAATAC 3198
408 GlnLeuAlaThrThrCysAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeu 427
3199 AAAGCAGCAGCAAGAAAGAAATCTTAATTTAGAATGCTTGCTAAATGAATGCATAGTCTT 3258
428 GluGluLysGluArgLysGluGluTyrThrLeuPheLysGlnLysLysAsnPhe 447
3259 TGTGAAATAGGAAATAAGTGGAA-----CAGCTAAAGGAAGCATTT 3303
448 GluArgGluArgSerPheThrGluAlaIleArgLeuGluLeuGluArgLysAla 467
3304 GCAAGGACACCAAGAAATTC-----TTAACAATAATTAGCA 3339
468 PheGluGluGluArgAlaSerTrpValLys-----GlnGlnPheLeu 481
3340 TTTCGTGAAGAAAGAAATCAGAAATCTGATGCTAGATTTGGAGACAGTGCAGCAAGCTCTG 3399
482 AsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPheSer 501
3400 AGATCTGAGATGACAGATAACCAACAATCTTAAG-----AGC 3438
502 GlySerSerAspProAsnLeuIleValHisSerArgProArgGlnLysLysLeuHis 521
3439 GAGGCTGGTGGTTAAAGCAAGAAATCATGACTTTAAAGGAAGAACAAACAATAATGCAA 3498
522 SerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAlaSer 541
3499 AAGGAATTAATGACTTATTACAGAGAATGAACGCTGATGAAGTAATG----- 3549
542 ProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIleSer 561
3550 -----AAGACTAAACATGAATG----- 3567
562 ValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThrAsp 581
3568 -----CAAAATCTAGAAATCAGAACCAATTTAGGAAC----- 3597
582 GlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCys 596
3598 -----TCTGTGAAGAAAGAGAGAGTGAAGAAATCAATGT 3633

RESULT 10

US-11-217-529-6026
; Sequence 6026, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIRO
; APPLICANT: KODAMA, YUKIO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIRO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6026
; LENGTH: 5658
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-6026

Alignment Scores: 6.58e-09 Length: 5658
Pred. No.:

Score: 205.00 Matches: 150
Percent Similarity: 37.6% Conservative: 118
Best Local Similarity: 21.1% Mismatches: 236
Query Match: 6.5% Indels: 208
DB: 7 Gaps: 35
US-10-644-084-2 (1-615) x US-11-217-529-6026 (1-5658)
QY 12 LeuCysThrGluAsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSerProSer 31
1366 TTAACGCTCAAAATGAAAACTGGCGGAATATGAAGATGAATCCACATA----- 1416
QY 32 SerLeuTyrSerGlnGlnVal---LeuCysSerSerVal----- 43
1417 ---TTGTCCAAACAACGCTCTGATTTATGTCGTCAAATTCATATCTCTTGATCACAAAT 1473
QY 44 -----ProLeuSerLysAsnValHisGlyValPheGlyVal 55
1474 TCAGTATCCAATGACTCTTAAGGAGCCTTTACCGCAA----- 1509
QY 56 PheCysThrGlyGluAsnIleGlu-----GlnSerIleSerTyrLeuAspGlnGluLeu 73
1510 -----GAGGAATCAAAATTTATTCAAAACATTTTACAAAACCGCATAGTGCT 1557
QY 74 ThrThrPheGlyPheProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGlu 93
1558 ACAACT-----GAATCAGACTCGCAAAAAATTTGTAACGGAA 1593
QY 94 LeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsn 113
1594 ---AGTTAGTTGAATTTAGAAAT-----ATTATTCAGTTACAGAGAAAAAT 1638
QY 114 LeuLeuAlaGlnGluSerValGluThrGlnAsnLeuLysLeuGlySerAspMetAspHis 133
1639 ---TCAGATCTTGGAGAGTAAACAGAAACTTA-----GCCGATAAA 1677
QY 134 LeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGlu----- 146
1678 TTAGAGTCAATGAAATAAATCTAAACAATATCTTAAGAAGATTGAAACCAAACTATA 1737
QY 147 ---ThrSerArgGluMetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLys 165
1738 AATGAGCTAAAGACGATCTCTTACCTTACAAAGCCAAAGACACGATTAGTAATCGAAA 1797
QY 166 AsnArgSerLeuHisGlnLeuLeuLysAsnGluLys-----Asp 178
1798 -----ATCCACGAGTTAGAAAGGAACGAGAAAAATTTAAGGACTGGACTTTGGAT 1848
QY 179 GluValGlnLysLeuGlnAsnIleIleAlaSerArgAlaThrGlnTyrAsnHisAspVal 198
1849 CAAAAGGCTTCACCTAATAACTCGATAATACAACTAATACTGAGACTTAAAGAGAGCTA 1908
QY 199 LysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsn 218
1909 -----GAATCCCAATCCAGACTTACAGCCCGCTATTCTCAAGCTTACCAGAGAA 1959
QY 219 LysLysAspLysAsnIleAlaMetAspValLeuAsn-----TyrVal 232
1960 TCTACCGAGAAT-----ATGTCACCTTTGAATAAGGAATTACAGGACCTATATGAT 2010
QY 233 GlyArgAlaAsp-----GlyLysArgGlySerTrpArg----- 243
2011 GGCAAAAGTAGCTATCAATTGAGCTAGGAGGAAAGTTCATCTAGGATATTGGCAGAA 2070
QY 244 -----ThrAspLysThrGluAlaArgGluAspGlu 254
2071 GAGAGTTCAATTTACTTTTCAATACGCTAGATTGACTTAAGCCGGAGAAC---GACCAA 2127
QY 255 MetTyrLysIleLeuLeuAsn-----AspTyrGluTyrArgGlnLys 268
2128 TTACGCAAGAGAGCAGTTAATTTACAAAATGCGCATCTCAAAAACAGATTCAAGAGCCAG 2187
QY 269 GlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMet 288

```
Db      2188 GAGACACTTAAGTACGTTCTTCTGCAAAATCGAAACTAAGTGTACTGAAACAGAAATTA 2247
Qy      289 IleserLeuLeuSerProGln-----LysLysLysProArgGlu 301
Db      2248 TCCAATTTGAAGTCAGACGACACTGAAATTCGAATTAGAAAAAACCCTAAGACAGAAA 2307
Qy      302 ArgAlaGluaspGlyThrValAlaIleserAspIleGluAspSerGlyGlu 321
Db      2308 -----CTAAGTGAACCTTCTCTGAGAAAAACAGT 2337
Qy      322 LeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsn 341
Db      2338 TTACATATATGTAACCTCAATTGCAAAACCTCCAAAAGGAACGTCGAGAACCTATTAGAT 2397
Qy      342 SerIleArgGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAla 361
Db      2398 GAATTTAAATCT-----TGTCAAAACAAATAGGTATATTAGAAATCGGCA 2448
Qy      362 SerLysValHisSerGluGlyLeuAsnGluAspValIleSer----- 376
Db      2449 AATGAACATAAAACCGAAGCTATTTCATAAAGCAACATACATTAAACAACTGGAAGAAGAT 2508
Qy      377 -----ArgGlnAspHisGlu 381
Db      2509 AATGACACAAAAGTTGAATGGTATCAGACAAAATTTGAAACTTTGAGGAAGATCATGAA 2568
Qy      382 Gln-----GluThrGluLysLeuGluLeuGluIle 391
Db      2569 TCTATTATGAATCTTTAAATGAAAAGCAAAATTTGGTTGAGAAATTCGCGTATGAAGTC 2628
Qy      392 GluArgCysLysGluMetIleLysAlaGlnGlnLeuLeuGln-----GlnGlnLeu 409
Db      2629 AAGCTTTGAAAAGGAATCGAAGAAAACAGATTCGTTTACATCTTACATCTACTA 2688
Qy      410 AlaThrThrCysAspAspThrThrSerLeuLeuArgAspCysTyrrLeuLeuGluGlu 429
Db      2689 GATGAATCTGTTAAACGATGAT-----TCCTTACGCAGAGAA-----CTGGAAGTCG 2736
Qy      430 LysGluArgLeuLysGluTrpThrLeuPheLysGluGlnLysLysAsnPheGluArg 449
Db      2737 AAAATAAACTTAACTGAGCTTATTCTCAATCCAGCAATACAAAAGCTTTTACGAGACT 2796
Qy      450 GluArgArgSerPheThrGluAlaAla----- 458
Db      2797 GCTGATAATCTCTACAGGAATGACATCATACAGATGAATCTAATAAATACATTTCT 2856
Qy      459 -----IleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpVal 476
Db      2857 AATCAAAATTCAAAGTTTAACTGATGAAAAGACCCACTTTTAGAGGATAAAGTTTCTCTG 2916
Qy      477 LysGlnGlnPheLeuAsnMetThrAsn---PheAspHisGlnAsnSerGluAsnValLys 495
Db      2917 AAAGAGCAATGCTCTAATTTGAATATGATGATTTTACAAAACCAAGCCATGGAANA 2976
Qy      496 LeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgPro 515
Db      2977 GAAAAATCCGAATTCAG-----AAAAAGATTCTTATTTTACAAAAT 3018
Qy      516 ArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThr 535
Db      3019 AACAAACAAAGAAATGGAAGCGGTCAAAATCTGAGTATGAGTCCAAAGTTGTGGAAGATTCAA 3078
Qy      536 LysSerLeuPro-----AlaSerProSerThrSerAspPheArg----- 548
Db      3079 AAGGATCTTGACAGCAAACTACATACGCTAATACTGCCCCAAAATAACTATGAACAGGAA 3138
Qy      549 ----GlnThrHisSerCysValSerGlu----- 556
Db      3139 TTGCAAAAACACGACAGATGTTTCCAAGACTATTAGTGAACATTAGAGAAACAATTACATACT 3198
Qy      557 ----HisSerSerIleSerValLeuAsnIleThrProGluGluSerLysProSerGluVal 575
```

```
Db      3199 TACAAGAGCTCAAGTAGAAACCTTTAAACCTAAACGCTGACCAG-----TTTGAAAAA 3249
Qy      576 AlaArgGluSerThrAspGlnLysTrpSerValGln 587
Db      3250 ACTCTAAAGGAAACGAAAAACAANTTGGAAATTCCCNA 3285
RESULT 11
US-11-217-529-76923
; Sequence 76923, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 76923
; LENGTH: 5787
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-76923
Alignment Scores:
Pred. No.: 1,41e-08 Length: 5787
Score: 201.50 Matches: 132
Percent Similarity: 37.9% Conservative: 97
Best Local Similarity: 21.9% Mismatches: 215
Query Match: 6.4% Indels: 160
DB: 7 Gaps: 25
US-10-644-084-2 (1-615) x US-11-217-529-76923 (1-5787)
Qy      60 GluAsnIleGluGlnSerIleSerTyrrLeuAspGlnLeuThrThrPheGlyPhePro 79
Db      2590 AAAAAATTCACGAGCAAAATTAATAAATCAAGAACGACCTTCAAGAAATGGAATCTAAG 2649
Qy      80 SerLeuTyrrGluGluSerLysSerLysGluAlaLysArgGluLeu---AsnIleValAla 98
Db      2650 AAGAAGTTTTTGGAGAAAAAGAACCAAAAACTGTCAATGAGTTGGAAAAATACCGAGAC 2709
Qy      99 ValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsn-----LeuLeu--- 115
Db      2710 TTGCTAAATCAGGAGAAAGAGAAATTTA-----AGAAAAAATGAGAGCTTATTAAT 2760
Qy      116 -----AlaGlnGluSerValGluThrGlnAsnLeuLysLeuGlySerAspMet 131
Db      2761 AGAGTTAAACACGCTCAGAAACATTAACAAAAACAATTCGATGACCTTTGTTCTGAAAG 2820
Qy      132 AspHisLeuGlnSerCysTyrrAlaLysLeuLysGluGlnLeuGluThrSerArgArgGlu 151
Db      2821 GATGAATATACGAGGAAAAAACTAGAGTTCCGCAAAATCTTTGAAGAGCTCATCAAAA 2880
Qy      152 MetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHis--- 170
Db      2881 ATCCAAGCCCTTCAAGAAACTATTAGAAACGGGAGCGACCTTAGAGAAATTCACACTCA 2940
Qy      171 -----GlnLeuLeuLysAsnGluLysAsp-----GluValGln 181
Db      2941 AAAAAATACGAACTAATAAAGCAAAATATCAGACTTGAACCTGTGATATTCTTAAGGAACAG 3000
Qy      182 LysLeuGlnAsnIleIleAlaSerArgAlaThrGlnTyrrAsnHisAspValLysArg--- 200
Db      3001 AGCTCCCGAGTCTGATCAAGAAATCTAAATGAGCTGGAAAAAGAAATAAGAGATTGA 3060
```

```
QY 201 -----LysGluArgGluTyrAsnLysLeuLysGluArgLeuHis--- 213
Db 3061 AAGATGTCATAAATCGAAGAGGAGAAATAAATCGTTCAACGATAGTTAAAGTTC 3120
QY 214 -----GlnLeuValMetAsnLysLysAspLysAsnIleAlaMet 226
Db 3121 TCAGAGAAGAGATCTTGACATAAAATTTGGTCACCTTTAGAGAAAAATTTGAACATTCGATG 3180
QY 227 AspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLys 246
Db 3181 TCAAGACTCAA-----TCCCTTGTAAACAGAAAAAT 3210
QY 247 ThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArg 266
Db 3211 TCAGATTGGTTCGAAATAATGAGAAATTTCAAG----- 3243
QY 267 GlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnMetLysLys 286
Db 3244 -----AAAGAAAAGGCAGCACTGAAT-----AACCAGTTGAAAAAT 3279
QY 287 GluMetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGlaAspGly 306
Db 3280 AATGAAAGTGAACCTGTAAAGATGAAGAAAAAATGATAATCACAAGAAAAAGAGCTTGCT 3339
QY 307 ThrGlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeuSerArgAspSer 326
Db 3340 ACTTTTCCAAGCAAGAGATGACGCGAGTAGTGAACATGGTAAAAATTTACTGCTGAA--- 3396
QY 327 ValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGln 346
Db 3397 -----CTAAAGACAGCAAGAAATCCAACCTTACCAG----- 3426
QY 347 TrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSer 366
Db 3427 -----TATAAATCTAATCAAAAAGATTAAAGAGAAATATTTCCAAC----- 3468
QY 367 GluGlyLeuAsnGluAspValIleSerArgGlnAspHisGluGlnGluThrGluLys 386
Db 3469 -----TTCCAAGAGAGAAACGAGGAAACAGGAGCAAAAGAAA 3504
QY 387 -----LeuGluLeuGluIleGluArg 393
Db 3505 AGAAACAGCCTGGTGGTCTCTGAACGATAGTAAATAATTAAAGAAATGGAA---GCTCGG 3561
QY 394 CysLysGluMetIleLysAlaGlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrCys 413
Db 3562 TTGTACAGGAAATATCCTTAATCAATCTTAACAAAAGAAATATCAGGAAATTTCTGTA 3621
QY 414 AspAspAspThrThrSerLeuLeuArgAspCysTyrLeu-----LeuGluGlu 429
Db 3622 GAAACGAATATTTCGCTACTAGACGTTCCACATCGTATTCAGATGATCCACTTGACAA 3681
QY 430 LysGluArgLeuLysGluGluTrpThrLeu-----PheLysGluGlnLysLysAsn 446
Db 3682 GAAGATATAATAAAAAAATACTATGACCTTCAATTTAGCGTTTACTGAAATAACTAGGAAT 3741
QY 447 PheGluArgGlu-----ArgArgSerPheThrGluAla 457
Db 3742 CTAGAGAATGAAATTGAAGAGAGAGAACTTAATTTCTAGATTGAGATTCTACTGAAACA 3801
QY 458 AlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLys 477
Db 3802 AGA-----CTAGCATCTTTCGCTTTTGGAGCAAAAAG-----ATTAA 3840
QY 478 GlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPhe 497
Db 3841 GCACAANTGAAGAAATTAATAAATAATGATCCAGGATATGAGCCCTAGTATTCTTTGGAC 3900
QY 498 SerAlaPheSer-----GlySerSerAspProAspAsnLeu 509
Db 3901 AGTATTTTAAATGAGCGCTAAATAACTGCTCCCTGACAAAGAGTCTGATATTACAAATTA 3960
QY 510 IleValHisSerArgProArgGlnLysLysLeu-----HisSer 522
Db 3961 ATGCTTGAGTGCATATTATTAAGAGACAATTTGGATATCGAAACAAGAGCTCTACTAGAT 4020
QY 523 ValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAlaSerPro 542
Db 4021 CGAGAAAATGCCATATCTCTTTACACAGATAAATTTAGAAAG----- 4062
QY 543 SerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIleSerVa 562
Db 4063 -----ATCCAAGAGAGAAAGCTCCCTGTCTCATCT 4089
QY 562 IleAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThrAspGln 582
Db 4090 TCTGATATTATTAACAACTGAAGTTCGAAGCCA-----GTGAAGAAAGAGTCAAAATCCTTG 4143
QY 582 nLysTrpSer 585
Db 4144 GAAGACAAGC 4153

RESULT 12
US-11-289-102-97
; Sequence 97, Application US/11289102
; Publication No. US2006012151A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Hyerim
; APPLICANT: Shaw, Peter M.
; APPLICANT: Clark, Edwin
; TITLE OF INVENTION: BIOMARKERS AND METHODS FOR DETERMINING SENSITIVITY TO
; MICROTUBULE-STABILIZING AGENTS
; FILE REFERENCE: 10338 NP
; CURRENT APPLICATION NUMBER: US/11/289,102
; PRIORITY FILING DATE: 2005-11-29
; PRIOR APPLICATION NUMBER: US 60/631,993
; NUMBER OF SEQ ID NOS: 395
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 97
; LENGTH: 4925
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-289-102-97

Alignment Scores:
Pred. No.: 8,15e-08 Length: 4925
Score: 192.00 Matches: 111
Percent Similarity: 40.5% Conservative: 112
Best Local Similarity: 20.2% Mismatches: 221
Query Match: 6.1% Indels: 107
DB: 7 Gaps: 18

US-10-644-084-2 (1-615) x US-11-289-102-97 (1-4925)
QY 11 ValLeuCysThrGluAsn-----LysAsnLeuSerGlnTyrThr 23
Db 1507 CTGGTATGCTTAAACAACACTGAGATTTTCAGAGAACAGCTCTGACCTCAGCCAGAAACTT 1566
QY 24 SerGluThrLysMetSerProSerSerLeuTyrSerGlnValLeuCysSerVal 43
Db 1567 AAAGAAACTCAGAGCAAAATACAGAGAGGC-TATGAAGAAAGTCTTGTAGTGTGCAAGAGCA 1625
QY 44 ProLeuSerLysAsnVal-----HisGlyValPheGlyValPheCysThrGly 59
Db 1626 GATGAACTCGGCTCTGTCTCACCTGAAAGCATGATAATTATTACATTTTCCAGAGCT 1685
QY 60 GluAsn-IleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPhePr 79
Db 1686 GAGGCTCAGGAAGAGAGAAATAAATGTCTAAAGCAGGATCTGCAGAAAT----- 1734
QY 79 oSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaVa 99
Db 1735 -GCATTAGAAGAAAGTGAAGAAATAAAGAAAGAAAGTGAAGAGTTA----- 1779
QY 99 IleuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSe 119
```



```
QY 111 ArgLysAsnLeuLeuAlaGlnGluSerValGluThrGlnAsnLeuLysLeuGlySer--- 129
Db 3214 ACAGATCATTACGTGTGAAATTCAGTGTAGTCCGAAATATATGGATGATGGTATGGC 3273
QY 130 -----AspMetAspHisLeuGln-----SerCysTyrAlaLysLeuLysGlu 143
Db 3274 GAAAGTAACGGTGATAAAGATATTATTAAGGCTAGTTTCATTATTTTCCAGTCTCGGTCAT 3333
QY 144 GlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArgGlnLeuGln 163
Db 3334 GAAAGGAATACCTCTGGAAACAAACTA-----ACTACC 3366
QY 164 CysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGlnLysAspGluValGlnLysLeu 183
Db 3367 TGTAG---AGAGATTTA---GCATTCGAAAGCAAAAGAAATGCCAATTTGGAAAGACC 3420
QY 184 GlnAsnIleIleAsnArgAlaThrGlnTyrAsnHisAspValLysArgLysGlu--- 202
Db 3421 GTAAATGATATG-----CAGAAAACTCACCCCATATATCGCGCAGGATGTC 3465
QY 203 -----ArgGluTyrAsnLysLeuLysGlnArgLeuHisGlnLeu 215
Db 3466 CAATGTTCTACCGAAATATTATGACGAGTTTGAGGATATTATGAAGAAATAAATACTCAAGTT 3525
QY 216 ValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrValGlyA-gAla 235
Db 3526 AACATCTTAAGGAACAAACACCATTTTCGAGAAGCGCTTGAGAGAGTT----- 3576
QY 236 AspGlyLysArgGlySerTrpArgTrpAspLysThrGluAlaArgAsnGluAspGluMet 255
Db 3577 -----ACAGAAAGAAATGNAGCA-----ATT 3597
QY 256 TyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAla 275
Db 3598 TATAAA-----GAGTACACTAATTTGCAACATGAACATATCTCGACTACAAGGT 3645
QY 276 GluLeu-----LysLysValLeu----- 281
Db 3646 GAACCTTGCAACGAAAGCAAACTTCTGTTAATGCTTAATAAGGTTTTAGTCTATGAA 3705
QY 282 -----GlnGlnMetLysLysGluMetIleSerLeuLysSerProGlnLysLysLys 298
Db 3706 AGTGAATTTGAACATGGAACAGAGGTTTCACAGCTTATCTCAACACAGAAAGAGGCA 3765
QY 299 ProArgGluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAsp 318
Db 3766 CACAAAGAAGAACTGAAAAG-----CTATTTAATGAGATTAGCGACATG 3810
QY 319 SerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGln 338
Db 3811 AGCGTAAACTT-----CTCAATGCCCCAGAAATGCAAAATGCTGAC 3849
QY 339 LeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAsp 358
Db 3850 TTGAATGATAAATTCAT-----AGGTTAAAGAAACACAGCGCAGCAAAAGTTGGAC 3900
QY 359 AsnGlnAlaSerLysValHisSerGluGlyLeu---AsnGluGluAspValIleSerArg 377
Db 3901 -----GCTTCCAGAAACAAACAACTGCTTGAACGAACTTGATGAATTTAAAGAT 3954
QY 378 GlnAspHisGlu-----GlnGluThrGlnLysLeuGluLeuGlu 390
Db 3955 ATGAGAAATGAATTTGGAAGAGAGTTTGGCTAGTGAAGAAATCTAAAGTCTTGAATGGAA 4014
QY 391 IleGluArgCysLysGluMetIleLysAlaGlnGlnGlnLeuLeuGlnGlnLeuAla 410
Db 4015 GCGAACTTAAGAGCATGAGTACAGGACGAGAGGCGTTTCAAAAATCAA----- 4065
QY 411 ThrThrCysAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLys 430
Db 4066 -----GAAGACGATACTTCCAAT-----CCATTAAATCGAAGAAATT 4101
QY 431 GluArgLeuLysGluGluTrpThrLeuPheLysGluGlnLysAsnPheGlu----- 448
```

```
Db 4102 CAATTACTTAAGAGGGAACCTTCAAGTGTTTAAAGAAACCAAGTGACTCTTCTGACATTATT 4161
QY 449 ArgGluArgArgSerPheThrGluAlaAlaIleArgLeuGlyLeuAlaIleAlaPhe 468
Db 4162 GAGAGAATGAAGAAATTTATGGAGGCGAGAGGCGCAAGACCATTTGANGAAGGCAACT 4221
QY 469 GluGluGluArg----- 472
Db 4222 GAGTTTGAAGGAAACCTAGAGAAATTTACCGGAAAAAATAGCGAGGAGGAGTAGACAAC 4281
QY 473 -----AlaSerTrpValLysGlnGlnPheLeuAsnMetThr 484
Db 4282 GGTGAAATATTGAGGAACCTGAAGAAACAATGCTCAAGCAA----- 4323
QY 485 AsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPheSerGlySerSer 504
Db 4324 ---TATGAAGAGAAACACGCTGAGAACAAATAA-----GAA 4356
QY 505 AspProAspAsnLeuIleValHisSerArg---ProArgGlnLysLysLeuHisSerVal 523
Db 4357 GCGGAAGAGAACTTTAAAGAAAAAGAAATCAGGTTACCAAGCGCAAGAAAGAAATTCAAAAATC 4416
QY 524 AlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAlaSerProSer 543
Db 4417 -----ATTAGCAACGCTAAGGAAGAATTG----- 4440
QY 544 ThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerIleSerValLeu 563
Db 4441 GAACAAGAGTTCCAAAGGAAG-----TTAGAAGAGAAATTCAGTTGCTTTACATTG 4491
QY 564 AsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThrAspGlnLys 583
Db 4492 TCT-----AGCAACAAAAAGGAAACGAAATTAAGTTGAA-----GACGAACATA 4536
QY 584 TrpSerValGlnSerArgProSerSerArgGlu 594
Db 4537 TGAATAGTCCCTCGAAGGGGAATTCGGAGAAG 4569
```

RESULT 14

```
US-10-953-349-20044
; Sequence 20044, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 20044
; LENGTH: 1648
; TYPE: DNA
; ORGANISM: Glycine max
US-10-953-349-20044
```

```
Alignment Scores:
Pred. No.: 4,08e-08 Length: 1648
Score: 188.00 Matches: 118
Percent Similarity: 38.3% Conservative: 100
Best Local Similarity: 20.7% Mismatches: 177
Query Match: 5.9% Indels: 174
DB: 6 Gaps: 26
```

US-10-644-084-2 (1-615) x US-10-953-349-20044 (1-1648)

```
QY 16 AsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSer 35
Db 190 AACGCTGAACACTCAGTCAGGAGTTTCAAAAGATGGAAGGATAAACCACTCTTTATGAG 249
QY 36 GlnGlnValLeuCysSerSerValProLeuSerLysAsnValHisGlyValPheGlyVal 55
```


Db 897 ---CCTCCAGCTTCTTATGTAGACTGCGAGTTCCTTCAAAGAAATACAGAAATTAT--- 950
QY 54 GlyValPheCysThrGlyGluAenIleGluGlnSerIleSerTyrLeuAenGlnGluLeu 73
Db 951 ---GGGAGATTCTCGAGATGTCAGTCACTTAT---GAAAGAGAGATC 992
QY 74 ThrThrPheGlyPhe-----ProSerLeuTyrGluGluSerLysSer----- 87
Db 993 ACAGCAGAGGGTGTGAGAGGCCAGAAATTTGCTCAACTTGGTCTTCGGCAGGCATTTCC 1052
QY 88 -----LysGluAlaLysArgGluLeuAenIleValAlaValLeuAenCys---MetAen 104
Db 1053 TGGAGGAGTGAACATGTCGGAG-----AACTGTGATGCTT 1091
QY 105 GluLeuValLeuGlnArgLysAenLeuLeuAlaGlnGlnSerValGluThrGlnAen 124
Db 1092 GACTGG-----GAGCAAGTCTGAAGCTTACAA 1121
QY 125 LeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGln 144
Db 1122 ---CCTGTCAAAGAGACATG---GCTTTAAATGAAGTCTTCGAGAAATTAATA 1169
QY 145 LeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCys 164
Db 1170 ---CATACTAACGAAGCAGGAGTGAATCCAAAGAA-----CTCCAGTGT 1214
QY 165 LysAenArgSerLeuHisGlnLeuLysAenGluLysAenGluValGlnLysLeuGln 184
Db 1215 AGTAACCTGATTAGAGAAGGGTTAAAGAACTACAGATGAAGATTACCAACAGCAA 1274
QY 185 AenIleAlaSerArgAlaThrGlnTyrAenHisAspValLysArgLysGluArgGlu 204
Db 1275 GTGTTCATT-----GATGTC----- 1289
QY 205 TyrAenLysLeuLysGluArgLeuHisGlnLeuValMetAenLys-----Lys 220
Db 1290 ATCAATAGCTTAAGAGAGATGTGAGAAATTAATTAAGACAAATACAAATAATCCTA 1349
QY 221 AspLysAenIleAlaMetAspValLeuAenTyrValGlyArgAlaAspGlyLysArgGly 240
Db 1350 GAGAAGAAAT-----GATACTAAAGACATTCGAGAATTTGGAAGAGGTTTAGCT 1400
QY 241 SerTrpArgThrAspLysThrGluAlaArgAenGluAspGluMet-----Tyr 256
Db 1401 AACACGCAAAACATCTTCAGGAATCCAGGAATGACAGGAATTTACAGCTTCAATTT 1460
QY 257 LysIleLeuAenAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAenAlaGlu 276
Db 1461 AAGAAGATCAAGGCTAATTTATGTGTGTTTACAGGAAGGTACATGACTGAAATGCAACAA 1520
QY 277 LeuLysLysValLeuGlnGln-----MetLysLysGlu 287
Db 1521 AAAAATAAATCTGTAAGTCAGTATTATTAGATGGACAAACCTTAAGCAAGAAGAGAA 1580
QY 288 MetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGlyThr 307
Db 1581 GAGGTAGAGAGACTACAA---CAACTCAAAAAGAACTGGAAGAGGCC----- 1625
QY 308 GlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeuSerArgAspSerVal 327
Db 1626 ---ACAGCTTCTGCTTGGACTTGTGTAACCGGGAAAAAGAGGCCCAAGAAACAGAGTTC 1682
QY 328 TrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAenSerIleArgLysGlnTrp 347
Db 1683 -----TTGCTTTACAGGAGGAATTCAGAACTTGAAGAGAAACCTGGAAGAAAGA 1736
QY 348 ArgIleLeuLysSerHisValGluLysLeuAenGlnAlaSerLysValHisSerGlu 367
Db 1737 CAGAAACTGAATCTAGACTTGAGAAATTCCTCACTCAAGTTAGAAATTTGCAATTTATG 1796
QY 368 GlyLeuAenGluLeuAsp-----ValIleSerArgGlnAspHisGlu-----Gln 382

Search completed: June 14, 2006, 04:09:52

Job time : 290 sec

Db 1797 TCTGAAATGTAAAGAACGAGAAATATATAAACTTTCAGCAGCAATCAACGAGTAAATAAT 1856
QY 383 GluThrGluLysLeuGluLeuGluIleGluArgCysLysGlu-----MetIleLys 399
Db 1857 GAGAAATGCAAAACTTAAACAGCAGGTTGCAAGGAGTGAAGAGCAAAATATTATGTCCTTAA 1916
QY 400 AlaGlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspThrThr 418
Db 1917 TTTGAGACAGCTCAGTTAAAGGATCAATTTAGAGGAAGTCTTGAAGTCAGATATTACCAAG 1976
QY 419 -----SerLeuLeuArgAspCysTyrLeuLeuGluGluLysGlu 431
Db 1977 GATACAAAAACACACATCTCTATCTGCTCCCGGATTCGCCCTTGTGAA-----GAG 2030
QY 432 ArgLeuLysGluGluTrpThrLeuPheLysGluLysGlnLysAenPheGluArgGluArg 451
Db 2031 AGGCTGAATCCT-----GCAGATATAAAAGAGAGCTTCTCAGCTGGCTCC 2075
QY 452 ArgSerPheThrGluAlaAlaIleArgLeuGlyLeu----- 463
Db 2076 AAAATGCACAGCTTCTTGCTCTGATGTGGAGCTTCTCACATGCCAGGACATCATCAAT 2135
QY 464 ---GluArgLysAlaPheGluGlu---GluArgAlaSerTrpValLysGlnPheLeu 481
Db 2136 TCTGATGCTGAACATTTCAAGAGAGAGTGAAGAGGTTAGTATATATGCTGCAAAACTG 2195
QY 482 AenMetThrAenPheAspHisGlnAenSerGluAenValLysLeuPheSerAlaPheSer 501
Db 2196 AAGAGCTCCATCTTAAAAAGAAACTTTAGAT----- 2228
QY 502 GlySerSerAspProAspAenLeuIleValHisSerArg----- 514
Db 2229 -----GAAGAGCTACTGAAACATAAAGATAGAAATCACAACCTTTAGAGAG 2273
QY 515 -----ProArgGlnLysLysLeuHisSerValAlaAenGlyValProAlaCysThrSer 532
Db 2274 TTAATTTGTAAGAAAAAGCATTTCAAGATCATGTCTATTAAAGTCTATAGACTGTGATTCA 2333
QY 533 LysLeuThrLysSerLeuProAlaSerProSer----- 543
Db 2334 GATGAAGCAAGAGATATCAGAGATGTACCTTACCCTTCTGGGAGCCAAACTGGATAAGTAC 2393
QY 544 -----ThrSerAspPheArgGlnThrHisSerCys 553
Db 2394 CACAGCTCTAAATGAGGAGCTTGATTCTTGTGTTACATCATATGAAGAAATCATTTGAATGT 2453
QY 554 ValSerGluHisSerSerIleSerValLeuAenIleThrProGluGluSer----- 570
Db 2454 GCTGACCAAGGCTTGCAATATCCACCTCCAGATTGCACATTTGGAAGAGAGAAATAA 2513
QY 571 -----LysProSerGluValAlaArgGluSerThrAspGlnLys 583
Db 2514 CATTTAGAGGATTTAATTTAGAAAGCCAGAGAAAAAGCCAGAAAAACCAAGATCAAAA--- 2570
QY 584 TrpSerValGlnSerArgProSerSer 592
Db 2571 ---AGCTTAGAAAAATCATCCGAAGTCC 2594

THIS PAGE LEFT BLANK